(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42323 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/EP01/13604

(22) International Filing Date:
22 November 2001 (22.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

0028495.0	22 November 2000 (22.11.2000)	GB
60/252,790	22 November 2000 (22.11.2000)	US

(71) Applicant *(for all designated States except US)*: **DEVGEN** NV [BE/BE]; Technologiepark 9, B-9052 Gent-Zwijnaarde (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NYS, Guy** [BE/BE]; Technologiepark 9, B-9052 Gent-Zwijnaarde (BE). **PLAETINCK, Geert** [BE/BE]; Technologiepark 9, B-9052 Gent-Zwijnaarde (BE). **BOGAERT, Thierry** [BE/BE]; Technologiepark 9, B-9052 Gent-Zwijnaarde (BE).

(74) Agents: **BOULT WADE TENNANT** et al.; Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/42323 A2

(54) Title: FLUORESCENT PROTEINS

(57) Abstract: There is disclosed an isolated nucleic acid molecule encoding a new florescent protein which is capable of emitting fluorescence upon irradiation by incident light, wherein said maximal absorbance of incident light is in the range of 440-480mm, and maximal fluorescence emission is in the range of 470-510mm.

FLUORESCENT PROTEINS

The present invention is concerned with fluorescent proteins and, in particular, with nucleic acid sequences encoding novel fluorescing proteins which have been isolated from coral species.

Fluorescent proteins, such as, green fluorescent protein from the luminescent jelly fish *Aequorea victoria* are extremely useful molecules by virtue of their ability to function as markers for gene expression and protein localisation within living cells. Fluorescent proteins can be produced *in vivo* by biological systems and can therefore be used to monitor and trace the progress of intracellular events.

In the present invention, the inventors have surprisingly identified completely novel fluorescing proteins from the coral species Anthozoa which have been sequenced and which can be used for *in vivo* labelling studies.

Therefore, according to a first aspect of the invention there is provided an isolated nucleic acid molecule encoding a fluorescent protein comprising an amino acid sequence illustrated in any of the polypeptide sequences of figures 3(a) to 3(d). The present inventors have advantageously identified 4 distinct nucleic acid molecules encoding fluorescing proteins which heretofore have not yet been described. In a further aspect, the invention comprises an isolated nucleic acid molecule encoding a protein capable of emitting fluorescence upon irradiation by incident light, wherein said maximal absorbance of said incident light is in the range 440-480 nm, in

particular 450-475 (maximum of excitation) and maximal fluorescence emission is in the range 470-510 nm, in particular 480-500 nm (maximum of emission).

According to the invention, at least 4 different fluorescent proteins (and nucleic acid sequences encoding said proteins) were obtained from species of coral, and in particular from species of coral belonging to the genus *Discosoma* and the genus *Polythoa*.

In addition, as can be seen from the data given hereinbelow, hybrids of fluorescent proteins derived from two or more different species from the genus *Polythoa* and/or *Discosoma* may also be used. Such hybrid fluorescent proteins of the invention may be obtained by suitable expression of hybrid (e.g.chimeric) nucleic acid sequences encoding such hybrid proteins, which in turn may for instance be obtained by suitably combining (two or more parts of) two or more naturally occurring nucleic acids (i.e. cDNAs and/or genes) encoding (native) fluorescent proteins, at least one of which has been obtained from a coral of the species *Polythoa* and/or *Discosoma* (and/or from another coral). This can be carried out by techniques known per se and/or as further described below, including but not limited to "gene shuffling" techniques.

A listing of the clones used in the invention is given in Figure 2. Also, an alignment of some of the clones used herein is given in Figure 8B.

The excitation- and emission-spectra for some of these proteins are given in the Figures, and are also summarized below:

Clone	Source	Mutations	Exitation max (nm)	Emission max (nm)
pGR7	*Polythoa* spec.	Q135R	469 (452)	490
pGR3	*Polythoa* spec.	N41D, 3'end	469 (452,489)	496
pGR13	*Polythoa* spec	none	469 (452)	490
pGR15	Hybrid	none	451 (440)	484

Accordingly, in one embodiment, the invention relates to a fluorescent protein with an emission spectrum which has:

- a maximum of emission (fluorescence - measured following exitation at 469 nm) at between 491 and 501 nm, and in particular at about 496 nm;
and preferably one, and more preferably both, of the following:

- an emission at 480 nm (fluorescence - measured following exitation at 469 nm) of between 30 and 40 % of the emission at the maximum of emission;

- an emission at 525 nm (fluorescence - measured following exitation at 469 nm) of between 35 and 45 % of the emission at the maximum of emission;
and with an exitation spectrum which has:

- a maximum of absorbance (measured at emission at 490 nm) at between 464 and 474 nm, and in particular at about 469 nm; and at least any one, preferably at least any two, more preferably at least any three, and most preferably all four of the following:

- an absorbance at 452 nm (measured at emission at 490 nm) of between 59 and 69 % of the absorbance at the maximum of absorbance;

- an absorbance at 456 nm (measured at emission at 490 nm) of between 54 and 64 % of the absorbance at the maximum of absorbance;

- an absorbance at 486 nm (measured at emission at 490 nm) of between 42 and 52 % of the absorbance at the maximum of absorbance;

- an absorbance at 489 nm (measured at emission at 490 nm) of between 63 and 73 % of the absorbance at the maximum of absorbance.

In another embodiment, the invention relates to a fluorescent protein with an emission spectrum which has:

- a maximum of emission (fluorescence - measured following exitation at 469 nm) at between 485 and 495 nm, and in particular at about 490 nm,
and preferably one, and more preferably both, of the following:

- an emission at 480 nm (fluorescence - measured following exitation at 469 nm) of between 46 and 56 % of the emission at the maximum of emission;

- an emission at 525 nm (fluorescence - measured following exitation at 469 nm) of between 33 and 43 % of the emission at the maximum of emission;
and with an exitation spectrum which has:

- a maximum of absorbance (measured at emission at 490 nm) at between 464 and 474 nm, and in particular at about 469 nm; and at least any one, preferably at least any two, more preferably at least any three, and most preferably all four of the following:

- an absorbance at 440 nm (measured at emission at 490 nm) of between 48 and 58 % of the absorbance at the maximum of absorbance;

- an absorbance at 452 nm (measured at emission at 490 nm) of between 55 and 65 % of the absorbance at the maximum of absorbance;

- an absorbance at 456 nm (measured at emission at 490 nm) of between 52 and 62 % of the absorbance at the maximum of absorbance;

- an absorbance at 480 nm (measured at emission at 490 nm) of between 48 and 58 % of the absorbance at the maximum of absorbance.

In yet another embodiment, the invention relates to a fluorescent protein with an emission spectrum which has:

- a maximum of emission (fluorescence - measured following exitation at 451 nm) at between 479 and 489 nm, and in particular at about 484 nm, and preferably one, and more preferably both, of the following:

- an emission at 470 nm (fluorescence - measured following exitation at 451 nm) of between 39 and 49 % of the emission at the maximum of emission;

- an emission at 525 nm (fluorescence - measured following exitation at 451 nm) of between 31 and 41 % of the emission at the maximum of emission;
and with an exitation spectrum which has:

- a maximum of absorbance (measured at emission at 484 nm) at between 446 and 456 nm, and in particular at about 451 nm; and at least any one, preferably at least any two, more preferably at least any three, and

most preferably all four of the following:

- an absorbance at 420 nm (measured at emission at 484 nm) of between 61 and 71 % of the absorbance at the maximum of absorbance;
- an absorbance at 440 nm (measured at emission at 484 nm) of between 86 and 96 % of the absorbance at the maximum of absorbance;
- an absorbance at 447 nm (measured at emission at 484 nm) of between 84 and 94 % of the absorbance at the maximum of absorbance;
- an absorbance at 470 nm (measured at emission at 484 nm) of between 61 and 71 % of the absorbance at the maximum of absorbance.

Also, any protein with an emission and/or exitation spectrum as indicated above preferably has a degree of sequence identity with at least one of the proteins encoded by the nucleic acid sequences shown in Figure 1, of at least 70%, preferably at least 80%, more preferably at least 90% and even more preferably at least 95% sequence identity with at least one of the proteins encoded by at least one of the nucleotide sequences depicted in Figure 1, in which the percentage sequence homology is determined as described hereinbelow.

For the some of the clones described hereinbelow, pertinent values are given in Figure 28.

Preferably, the nucleic acid molecule is a DNA and more preferably a cDNA molecule. The cDNA molecules are preferably isolated from the Discosoma or Polythoa

genus of coral although they may also be synthetically prepared using techniques which would be well known to practitioners skilled in the art. Preferably, the nucleic acid sequences encoding the novel proteins are as set forth in Figure 1.

Preferably, the nucleic acid molecule is substantially homologous to the nucleic acid sequences depicted in Figure 1. Even more preferably the nucleic acid molecule has at least 70, preferably at least 80, more preferably at least 90 and even more preferably at least 95% sequence identity to at least one of the nucleic acid sequences depicted in Figures 1 and even more preferably comprises any of the nucleic acid sequences of Figure 1.

The fluorescent proteins of the invention can be used for any application known per se for fluorescent proteins described in the art, such as for the green fluorescent protein from *Aequorea victoria* mentioned above. Such applications will be clear to the skilled person, and may include, but are not limited to, the applications of such "GFPs" mentioned in the relevant prior art, such as WO 95/07463, WO 97/11094, WO 97/42320, WO 98/06737 and WO 97/41228.

As such, the fluorescent proteins of the invention (and/or the nucleic acid sequences encoding these proteins) may be used as a label and/or marker, and in particular as a genetic marker and/or an expression marker, for instance in the fields of (micro-)biology, biochemistry and/or molecular biology.

For example, the fluorescent proteins of the

inventions (and/or the nucleic acid sequences encoding these proteins) may be used in *in vitro* applications, such as hybrisation assays and/or immunological assays (e.g. ELISA's).

However, fluorescent proteins of the invention are particularly suited for applications *in vivo*, including but not limited to expression and/or use in bacteria, protozoa, fungi, algi, yeast cells or other micro-organisms; in (cells or tissues of) plants and/or animals; and/or in cells or cell lines derived from plant cells or animal cells.

One particularly preferred application involves the expression and use in species of nematode, such as *C.elegans*, e.g. for screens or assays involving the use of such nematodes.

Some other possible applications include, but are not limited to:

- follow up of a protein tagged with a fluorescent protein during the purification of said protein (e.g. using chromatography techniques);

- *in vivo* expression analysis;

- investigation of the transport of proteins etc. across biological membranes; and/or (other) qualitative and/or quantitative detection techniques and/or analytical techniques.

The nucleic acid molecules of the present invention are particularly useful in processes for labelling

polypeptides of interest, e.g., by the construction of genes encoding fluorescent fusion proteins. Fluorescence labelling via gene fusion is site-specific and eliminates the present need to purify the labelled proteins *in vitro* and microinject them into cells. Sequences encoding the fluorescing proteins of the present invention may be used for a wide variety of purposes as are well known to those working in the field. For example, the sequences may be employed as reporter genes for monitoring the expression of the sequence fused thereto; unlike other reporter genes, the sequences require neither substrates nor cell disruption to evaluate whether expression has been achieved. Similarly, the sequences of the present invention may be used as a means to trace lineage of a gene fused thereto during the development of a cell or organism. Further, the sequences of the present invention may be used as a genetic marker; cells or organisms labelled in this manner can be selected by e.g. fluorescence-activated cell sorting. The sequences of the present invention may also be used as a fluorescent tag to monitor protein expression *in vivo* and/or *in vitro* or to encode donors or acceptors for fluorescence resonance energy transfer. Other uses for the sequences of the present invention would be readily apparent to those working in the field, as would appropriate techniques for fusing a gene of interest to an oligonucleotide sequence of the present invention in the proper reading frame and in a suitable expression vector so as to achieve expression of the combined sequence.

Similarly fusion proteins including an antibody fused to the fluorescing protein may also be generated for

in vivo labelling, for example. In such an embodiment the nucleic acid molecule of the invention encoding the fluorescing protein will be operably linked to the sequence encoding the antibody. As would be well known in the art only a small portion of an antibody molecule, the paratope, is involved in binding to the epitope of a protein and a nucleic acid molecule encoding the paratope may be used to generate a labelled molecule specific for the paratope of interest.

A fusion protein of the 3' sequence of Discosoma coupled to the 5' sequence of Polythoa 2 was also generated using the nucleic acid sequences encoding the Polythoa 2 and Discosoma 1 protein, for expression in a prokaryotic and eukaryotic expression system, which protein sequences are illustrated in Figure 7. The plasmid pGR15 encoding the sequence of the Polythoa 2-Discosoma 1 hybrid was the vector used for expression of the fusion protein in *E.coli*, whereas plasmid pGR18 was utilised for eukaryotic expression in COS cells. Plasmid pGR20 was used for expression in *C.elegans* and transformation of the relevant cells or organism using these vectors resulted in expression of a fluorescing protein.

As outlined in more detail in the examples below, mutant or hybrid proteins were also developed to investigate their absorbance and emission spectra compared to the wild type Polythoa and Discosoma proteins. The proteins and polypeptides encoded by plasmids pGR3 and pGR7 described herein contain a 109 thioredoxin associated fragment in fusion with the Polythoa 2 fluorescing protein. Furthermore, plasmid

pGR7 encodes a protein with the mutation Q136R while a further plasmid pGR10 expresses a I106T mutant.

An antisense molecule capable of hybridising to the nucleic acid molecules of the invention under conditions of high stringency also forms part of the invention.

Stringency of hybridisation as used herein refers to conditions under which polynucleic acids are stable. The stability of hybrids is reflected in the melting temperature (Tm) of the hybrids. Tm can be approximated by the formula:

$$81.5°C+16.6(\log_{10}[Na^+]+0.41\ (\%G\&C)-600/1$$

wherein 1 is the length of the hybrids in nucleotides. Tm decreases approximately by 1-1.5°C with every 1% decrease in sequence homology.

The term "stringency" refers to the hybridisation conditions wherein a single-stranded nucleic acid joins with a complementary strand when the purine or pyrimidine bases therein pair with their corresponding base by hydrogen bonding. High stringency conditions favour homologous base pairing whereas low stringency conditions favour non-homologous base pairing.

"Low stringency" conditions comprise, for example, a temperature of about 37°C or less, a formamide concentration of less than about 50%, and a moderate to low salt (SSC) concentration; or, alternatively, a temperature of about 50°C or less, and a moderate to high salt (SSPE) concentration, for example 1M NaCl.

"High stringency" conditions comprise, for example, a temperature of about 42°C or less, a formamide concentration of less than about 20%, and a low salt (SSC) concentration; or, alternatively, a temperature of about 65°C, or less, and a low salt (SSPE) concentration. For example, high stringency conditions comprise hybridization in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C (Ausubel, F.M. *et al.* Current Protocols in Molecular Biology, Vol. I, 1989; Green Inc. New York, at 2.10.3).

"SSC" comprises a hybridization and wash solution. A stock 20X SSC solution contains 3M sodium chloride, 0.3M sodium citrate, pH 7.0.

"SSPE" comprises a hybridization and wash solution. A 1X SSPE solution contains 180 mM NaCl, 9mM Na_2HPO_4 and 1 mM EDTA, pH 7.4.

However, other conditions and reagents also result in stringent hybridisation conditions and these are generally well known to the skilled practitioner (Molecular Cloning A Laboratory Manual, J. Sambrook *et al.*, Cold Spring Harbour Press, 1989, or Current Protocols in Molecular Biology, F.M. Ansubel, *et al.*, eds., John Wiley & Sons Inc., New York.

As would be appreciated by those skilled in the art, the presence of introns in a nucleic acid sequence can lead to enhanced expression levels. One of the preferred nucleic acid molecules of the invention, the sequence of which is depicted in Figure 2(b), includes a synthetic intron in addition to a 5' UTR including a

Kozak site.

Fluorescent proteins or functional equivalents, fragments or variants thereof encoded by the nucleic acid molecules of the invention also form part of the invention. Furthermore, according to an even further aspect, the invention comprises an isolated fluorescent protein capable of emitting fluorescence upon irradiation by incident light wherein the maximal absorbance of said incident light is in the range 440-480 nm, in particular 450-475 nm (maximum of excitation) and maximal fluorescence emission is in the range 470-510 nm, in particular 480-500 nm (maximum of emission). The invention also comprises an isolated fluorescent protein comprising an amino acid sequence which has at least 70, preferably at least 80, more preferably at least 90 and even more preferably at least 95% sequence identity to the amino acid sequence depicted in any of Figures 3 to 8.

Functional equivalents, fragments or variants of the polypeptide of the invention are those molecules that retain the distinct fluorescing capability of the polypeptides of the invention.

The DNA molecules according to the invention may, advantageously, be included in a suitable expression vector to express the fluorescent protein encoded therefrom in a suitable host. Incorporation of cloned DNA into a suitable expression vector for subsequent transformation of said cell and subsequent selection of the transformed cells is well known to those skilled in the art as provided in Sambrook *et al.* (1989), Molecular Cloning, A Laboratory Manual, Cold

Spring Harbour Laboratory Press.

An expression vector according to the invention includes a vector comprising a nucleic acid according to the invention operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. The term "operably linked" refers to a juxta position wherein the components described are in a relationship permitting them to function in their intended manner. Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide according to the invention. Thus, in a further aspect, the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell, transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, and optionally a promoter for the expression of said nucleotide sequence and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, ampicillin resistance.

The precise nature of the regulatory sequences needed for expression of the fluorescing protein can vary between species or cell types. They will, however, generally include 5' non-transcribing and 5' non-

translating sequences involved in initiation or regulation of transcription and translation respectively. Regulatory elements required for expression generally include promoter sequences to bind RNA polymerase and transcription initiation sequences for ribosome binding. For example, a bacterial expression vector may include a promoter such as the lac promoter and for translation initiation the Shine-Dalgarno sequence and the start codon AUG. Similarly, a eukaryotic expression vector may include a heterologous or homologous promoter for RNA polymerase II, a downstream polyadenylation signal, the start codon AUG, and a termination codon for detachment of the ribosome. Such vectors may be obtained commercially or assembled from the sequences described by methods well known in the art.

Nucleic acid molecules according to the invention may be inserted into the vectors described in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense nucleic acids may be produced by synthetic means.

As discussed in the examples provided it is desirable to enhance the performance or expression levels of the fluorescent proteins in organisms or cells other than those from the coral species from which the proteins or polypeptides of the invention are derived. Every organism adopts a preferred codon usage which is related to the presence and expression of tRNA genes and which involves post-transcriptional expression regulation. Such optimal codon usage has been determined for a number of organisms. In the present embodiment a vector was generated for optimal

expression in the nematode *C.elegans*. Therefore, when the host to be transfected with a vector including the nucleic acid molecules of the invention is *C.elegans*, the vector may comprise the plasmid pGR10, described in the example below, which includes the nucleotide sequence depicted in Figure 2(a).

Similarly, the introduction of synthetic introns can result in enhancements of expression levels. A preferred nucleic acid molecule including such a synthetic intron for increased expression levels in *C.elegans* is particularly preferred, which molecule is described in Figure 2(b).

Preferred vectors according to the invention comprise the plamsids designated pGR3, pGR4, pGR5, pGR6, pGR7 and pDW2700, the sequences of which are illustrated in Figures 9 to 14 respectively. Other preferred plasmids according to the invention comprise plasmids designated pGR1, pGR8, pGR13, pGR14, pGR15, pGR16, GR17, pGR18, pGR19, pGR20 and pGR10 identified in the example provided, and which would be readily producible by the skilled practitioner using the method steps described.

In accordance with the present invention, a defined nucleic acid includes not only the identical nucleic acid but also any minor base variations including in particular, substitutions in cases which result in a synonymous codon (a different codon specifying the same amino acid residue) due to the degenerate code in conservative amino acid substitutions. The term "nucleic acid sequence" also includes the complementary sequence to any single stranded sequence

given regarding base variations.

The present invention also advantageously provides nucleic acid sequences of at least approximately 10 contiguous nucleotides of a nucleic acid according to the invention and preferably from 10 to 50 nucleotides of the nucleic acid sequences set forth in Figures 1 and 2. These sequences may, advantageously be used as probes or primers to initiate replication, or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with the sample under hybridising conditions and detecting for the presence of any duplex or triplex formation between the probe and any nucleic acid in the sample.

Letters utilised in the sequences according to the invention which are not recognisable as letters of the genetic code signify a position in the nucleic acid sequence where one or more of bases A, G, C or T can occupy the nucleotide position. Representative letters used to identify the range of bases which can be used are as follows:

M: A or C
R: A or G
W: A or T
S: C or G
Y: C or T
K: G or T

V: A or C or G
H: A or C or T
D: A or G or T
B: C or G or T
N: G or A or T or C

According to the present invention, degenerate primers were utilised to fully identify the sequence of the nucleic acid encoding the proteins of the invention. Those novel molecules as described in the example provided also form part of the present invention.

According to the present invention these probes may be anchored to a solid support. Preferably, they are present on an array so that multiple probes can simultaneously hybridize to a single biological sample. The probes can be spotted onto the array or synthesised *in situ* on the array. (See Lockhart *et al.*, Nature Biotechnology, vol. 14, December 1996 "Expression monitoring by hybridisation to high density oligonucleotide arrays". A single array can contain more than 100, 500 or even 1,000 different probes in discrete locations.

The nucleic acid sequences, according to the invention may be produced using such recombinant or synthetic means, such as for example using PCR cloning mechanisms which generally involve making a pair of primers, which may be from approximately 10 to 50 nucleotides to a region of the gene which is desired to be cloned, bringing the primers into contact with mRNA, cDNA, or genomic DNA from a suitable biological source, and in particular from (a cell of) a species of coral, more particularly from (a cell of) a species

of coral from the genus Polythoa and/or the genus Discosoma, performing a polymerase chain reaction under conditions which brings about amplification of the desired region, isolating the amplified region or fragment and recovering the amplified DNA. Some of the primers suitable for the aforementioned method include, but are not limited to, the individual primers mentioned in Table 1 as well as the combinations thereof mentioned in Table 2. Generally, such techniques are well known in the art, such as described in Sambrook *et al.* (Molecular Cloning: a Laboratory Manual, 1989). Another suitable technique involves "gene shuffling" (DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution: Proc. Natl. Acad. Sci. Vol 91, pp 10747-10751, October 1994.

Therefore, it is also envisaged that - based upon the disclosure herein and (for instance) using one or more of the primers listed in Table 1 or a suitable combination thereof (including but not limited to the combinations mentioned in Table 2 - the skilled person will be able to isolate (nucleic acids encoding) additional fluorescent proteins of the invention from other suitable biological sources, and in particular from other species of coral such as (other) species from the genus *Polythoa* and/or the genus *Discosoma;* and such (nucleic acids encoding such) additional fluorescent proteins are also within the scope of the present invention.

In one preferred embodiment, such any nucleic acids will have at least 70%, preferably at least 80%, more preferably at least 90% and even more preferably at

least 95% sequence identity with at least one of the nucleotide sequences depicted in Figure 1, in which the percentage sequence homology is determined as described above; and/or is capable of hybridizing with at least one of the nucleotide sequences depicted in Figure 1 under conditions of high stringency, again as described above.

The term "homologous" describes the relationship between different nucleic acid molecules or amino acid sequences wherein said sequences or molecules are related by partial identity or similarity at one or more blocks or regions within said molecules or sequences. Homology may be determined by means of computer programs known in the art.

Substantial homology preferably carries with it that the nucleotide and amino acid sequences of the fluorescent proteins of the invention comprise a nucleotide and amino acid sequence fragment, respectively, corresponding and displaying a certain degree of sequence identity to the sequences set forth in Figures 1 and 2 for the nucleotide sequences and 3 to 8 for the polypeptide sequences. Preferably they share an identity of at least 30 %, preferably 40 %, more preferably 50 %, still more preferably 60 %, most preferably 70%, and particularly an identity of at least 80 %, preferably more than 90 % and still more preferably more than 95 % is desired with respect to the nucleotide or amino acid sequences depicted in Figures 1 to 8 respectively. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global

sequence alignment, can be determined using, for example, the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. 6 (1990), 237-245.) In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Further programs that can be used in order to determine homology/identity are described below and in the examples. The sequences that are homologous to the sequences described above are, for example, variations of said sequences which represent modifications having the same biological function, in particular encoding proteins with the same or substantially the same receptor specificity, e.g. binding specificity. They may be naturally occurring variations, such as sequences from other mammals, or mutations. These mutations may occur naturally or may be obtained by mutagenesis techniques. The allelic variations may be naturally occurring allelic variants as well as synthetically produced or genetically engineered variants. In a preferred embodiment the sequences are derived from a human.

A further aspect of the invention provides host cells transformed or transfected with a vector according to the invention. Such cells can be of prokaryotic or eukaryotic origin. Suitable prokaryotes include gram positive or negative organisms including *E.coli*, Bacillus spp, Pseudomonas spp, or salmonella typhimurium. The expression vector used to transform the prokaryotic cells, and particularly *E.coli*, preferably comprises plasmids designated pGR3 and pGR7, the sequences of which are illustrated in Figure

9 and 13 respectively. Eukaryotic organisms include yeasts or fungi and plant cells which utilise a transfection system based on infection by Agrobacterium tumefaciens.

The vectors can also be used to transform cells in tissue culture in addition to non-human organisms and these also form part of the invention. Typical mammalian tissue culture cells include COS-7, HEK-293, BHK, CHD, HELA cells and the like. Suitable organisms which may be useful to monitor expression of proteins using the novel fluorescing proteins of the invention include *C.elegans*, which is particularly advantageous as the fluorescing protein can be viewed *in vivo*.

When the organism to be transformed with the appropriate vector is *C.elegans*, the vector preferably comprises the sequence of the plasmid illustrated in Figure 12 or a vector adapted for expression of heterologous proteins in the *C.elegans* including the nucleotide sequences illustrated in Figure 2.

Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as *E.coli*, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method by procedures well known in the art. Alternatively, $MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell or by electroporation.

When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in lipsosomes, or virus vectors may be used. Eukaryotic cells can also be cotransfected with DNA sequences encoding the fusion polypeptide, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40(SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the proteins (*Eukaryotic Viral Vectors, Cold Spring Harbour Laboratory, Gluznan ed., 1982.*

Also encompassed within the scope of the present invention is a method of producing a polypeptide according to the invention comprising cultivating a host cell or tissue transformed or transfected with the appropriate vector of the invention under conditions suitable for expression of the flourescent protein and optionally recovering the expressed protein. The protein may be recovered and purified from the recombinant cell cultures by methods known in the art, including ammonium sulfate or ethanol precipitation acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxyapatite chromatography and lectin chromatography.

In a further aspect, the invention also comprises an oligonucleotide probe or primer, and which comprises a sequence that selectively hybridises to a nucleic acid

molecule according to the invention. The oligonucleotide preferably comprises a sequence of at least 10 contiguous nucleotides and is preferably between 10 and 50 nucleotides in length.

Advantageously, the novel proteins of the invention, as aforementioned, are particularly useful for monitoring expression of proteins within biological systems and the subcellular localisation or trafficking of proteins. To determine the expression pattern of a particular protein of interest it suffices in principle to make a fusion between the promoter of the gene of interest and the sequence encoding the fluorescing protein. Upon introduction of a vector with the promoter-fluorescent protein of the invention fusion into a cell or organism, any expression induced by the promoter can easily be monitored by following the expression of the protein of the invention. To monitor the subcellular expression of a protein it generally suffices to make a fusion between the protein of interest and the GFP protein, which can be done at either the N or C terminals of the protein.

Therefore, in a further aspect the present invention comprises a method for selecting cells capable of expressing a protein of interest, comprising introducing into said cells a vector comprising the nucleotide sequence of a fluorescent protein according to the invention operatively linked to a promoter or regulatory region of the protein of interest, cultivating the cell under conditions necessary for expressing the protein of interest and monitoring for any fluorescent following expression of said

fluorescent protein.

In accordance with the present invention, a protein of interest includes any protein to be monitored or labelled by virtue of being attached or expressed together with the proteins of the invention. The techniques for generating fusion proteins using the proteins of the invention are well known to those in the art.

A particular use of fluorescent proteins consists of the construction of a synthetic protein harboring a donor fluorescent protein and an acceptor fluorescent protein, connected with a binding protein moiety. The two fluorescent proteins change conformation upon binding of an analyte to the binding protein moiety. The binding protein moiety has an analyte-binding region which binds an analyte and causes the indicator to change conformation upon exposure to the analyte. The donor fluorescent protein is covalently coupled to the binding moiety. The acceptor fluorescent protein moiety is also covalently coupled to the binding protein moiety. In the fluorescent indicator the donor moiety and the acceptor moiety change position relative to each other when the analyte binds to the analyte binding region, altering fluorescence resonance energy transfer between the donor moiety and the acceptor moiety when the donor moiety is excited. Such a system has been described previously by Tsien et al. WO 98/40477 and Garman WO 94/28166. These molecules are very efficient in measuring internal concentrations of analytes such as cAMP, Ca^{2+}, etc. as for measurement of internal enzymatic activities of enzymes such as proteases, esterases, etc. The novel

fluorescent proteins according to the present invention and functional equivalents, derivatives or fragments thereof can be used to develop new FRET molecules.

Therefore, in a further aspect the present invention comprises a method for producing fluorescence resonance energy transfer comprising; providing an acceptor molecule comprising a fluorescent protein according to the invention providing an appropriate donor molecule for the fluorescent protein; and bringing the donor molecule and acceptor molecule into sufficiently close contact to allow fluorescence resonance energy transfer. Alternatively, the donor molecule can be the fluorescent protein of the invention in which case an appropriate acceptor molecule for the fluorescent protein is provided.

The invention may be more clearly understood from the following description of an exemplary embodiment with reference to the accompanying Figures wherein:

Figure 1 is an illustration of the nucleotide sequences encoding for fluorescent proteins from the Polythoa and Discosoma species of coral.

Figure 2 (a) is an illustration of the sequence of the DNA fragment encoding Polythoa 2 protein with optimal codon usage for expression in *C.elegans*.

(b) is an illustration of the sequence from (a) further including introns and a 5' untranslated region containing a Kozak

sequence.

Figure 3(a-d) is an illustration of the polypeptide sequences of Polythoa 1 and 2 and Discosoma 1 and 2 encoded by the nucleic acid molecules of the invention.

Figure 4 is an illustration of the sequence of a Polythoa fusion protein encoded by plasmid pGR3 and which includes a 109 amino acid thioredoxin fragment fused to the Polythoa 2 polypeptide sequence.

Figure 5 is an illustration of the sequence of a Polythoa 2 fluorescent fusion protein in pGR7 which also incorporates the 109 thioredoxin amino acid fragment.

Figure 6 is an alignment of the proteins encoded by the plasmids indicated A-J therein.

Figure 7 is a further alignment of the protein sequences of the Polythoa 2, Discosoma 1 hybrid and the proteins encoded by the plasmids indicated therein.

Figure 8 (a) is a further alignment of the translation products from the DNA fragments indicated therein.

Figure 8 (b) is an alignment of some of the clones used in the present invention.

Figure 9 is an illustration of the nucleotide sequence of plasmid pGR3.

Figure 10 is an illustration of the nucleotide sequence of plasmid pGR4.

Figure 11 is an illustration of the nucleotide sequence of plasmid pGR5.

Figure 12 is an illustration of the nucleotide sequence of plasmid pGR6.

Figure 13 is an illustration of the nucleotide sequence of plasmid pGR7.

Figure 14 is an illustration of the nucleotide sequence of plasmid pDW2700.

Figure 15 is a graphic representation of the emission spectrum of the thioredoxin-FP-fusion protein from pGR3 at (a) 452 nm and (b) 489 nm excitation.

Figure 16 is a graphic representation of the emission spectrum of thioredoxin-FP-fusion protein from pGR3 at 469 nm excitation.

Figure 17 is a graphic representation of the pGR3 excitation spectrum at an emission of 490 nm.

Figure 18 is a graphic representation of the excitation spectrum of thioredoxin-FP-

fusion protein from pGR7 at 490 nm emission.

Figure 19 is a graphic representation of the emission spectrum of thioredoxin-FP-fusion protein from pGR7 at 452 nm excitation.

Figure 20 illustrates combined emission and excitation spectra of thioredoxin-FP-fusion protein from pGR7.

Figure 21 is a graphic representation of the emission spectrum of thioredoxin-FP-fusion protein from pGR13 at 452 nm excitation.

Figure 22 is a graphic representation of the emission spectrum of thioredoxin-FP-fusion protein of pGR13 at 469 nm excitation.

Figure 23 is a graphic representation of the excitation spectrum of the thioredoxin-FP-Fusion proteins from pGR13 at 490 nm emission.

Figure 24 is a graphic representation of the emission spectrum of thioredoxin-FP-fusion protein pGR15 at (a) 489 nm excitation and (b) 451 nm excitation.

Figure 25 is a graphic representation of the emission spectrum of thioredoxin-FP-

fusion protein pGR15 at 440 nm excitation.

Figure 26 is a graphic representation of the emission spectrum of thioredoxin-FP-fusion protein pGF15 at 440 nm excitation.

Figure 27 is a list of the clones used in accordance with the invention.

Figure 28 is a list of pertinent absorbance and emission values for some of the clones used.

Examples:

1) Isolation of cDNA encoding for new fluorescent proteins

a) Isolation of RNA

Two brightly fluorescent Anthozoa species (Polythoa and Discosoma species) were used to isolate fluorescent proteins. This type of coral can be obtained from aquarium supply outlets, but such corals can be obtained from various coral reefs. The corals, and more particularly the polyps expressing high levels of fluorescent protein were flash-frozen in liquid nitrogen. Methods to isolate material samples are common in molecular biology techniques, and have been described in "Current Protocols in Molecular Biology", ed. by Ausubel et al., John Wiley & Sons, Inc.

Total RNA was isolated using TRIzol™ Reagent (Cat. NO. 15596; Life Technologies), according to the manufacturers procedure, from the frozen specimens and the total RNA was finally re-suspended in DEPC water (Current protocols in Molecular biology, ibid).

b) First strand cDNA synthesis

First strand cDNA was prepared using the total RNA isolations as described above from the Polythoa or the Discosoma species. Random primers were provided by Life Technologies (Cat. NO. 48190-11) and cDNA was synthesized using the Superscript II kit (Cat. NO. 18064-022; Life Technologies). The protocol to generate cDNA, by RT-PCR was performed according the instructions of the manufacturers.

c) PCR with degenerate primers:

To isolate full cDNA sequences encoding for new fluorescent proteins, a series of PCR procedures were performed using the cDNA isolated as described above. For these experiments, the following synthetic degenerate primers were used:

oGR1: CACCACATGGAAGGAWRYKTNRAYGG;

oGR2: ACCACATGGAAGGATGCKTNRAYGGNCA;

oGR3: AATTTGTGATCAAGGGCRARGGNRWNGG;

oGR4:GTGATCAAAGGTGGACCNYTNCCNTT;

oGR5:GACATATTGTCAACAGAGTTYMANTAYG;

oGR6:CATATTGTCAACAGAGTTYMANTAYGG;

oGR7:ATCCTGACGACATACCAGAYTAYHWNAA;

oGR8:GACTATTTCAAGCAGTCGTKYCCNGMNGG;

oGR9:CATGGGAAAGGTCCTTGCAYTWYGARGA;

oGR10:GGTGACATCTCCTTTCARNAYNCC;

oGR11:CATATTCTCAGTGGANGSNTCCCA;

oGR12:CACAGGTCCATCGSNAGGRAARTT;
oGR13:CCATCGGCAGGAAARTTNANNCC;
oGR14: TGAATACCCTGTTTCCRTANTKRAA

The first strand cDNAs as isolated above were subjected to PCR amplification using the set of degenerate primers (oGR1 till oGR14) and Amplitaq Gold (Perkin Elmer) as Polymerase. Concentrations, buffers were as provided by the manufacture or minor modifications were applied as known in the art.
The PCR conditions were as followed:
An initial denaturation step at 95°C for 10', followed by 25 cycles of touch down PCR (30" at 95°C, 1' at 55°C (-0.2°C/cycle) and 1' at 72°C) and followed by 15 cycles of PCR (95°C for 30", at 50°C for 1' and 72°C for 1'). The resulting PCR products were analyzed on standard agarose gel and the DNA fragments of interest were isolated and cloned into vector pCR-XL-TOPO vector (Cat. NO. K4700-20; Invitrogen).
Following primer combinations resulted in the isolation of appropriate DNA fragments
On Polythoa first strand cDNA:
oGR1/oGR14, oGR6/oGR11, oGR2/oGR11, oGR3/oGR11, oGR4/oGR11, oGR5/oGR11, oGR1/oGR11,
on Discosoma first strand cDNA:
oGR1/oGR10, oGR1/oGR11, oGR6/oGR10, oGR6/oGR11, oGR2/oGR11
oGR1/oGR12, oGR1/oGR14, oGR6/oGR12, oGR6/oGR13, oGR3/oGR11, oGR4/oGR11, oGR5/oGR11, oGR8/oGR11, oGR9/oGR11

It would be apparent to a person skilled in the art that other primer combinations will also result in the isolation of DNA fragments encoding for fluorescent proteins, such as the primer combinations. oGR1/oGR13,

oGR2/oGR10, oGR2/oGR12, oGR2/oGR13, oGR2/oGR14,
oGR3/oGR10, oGR3/oGR12, oGR3/oGR13, oGR3/oGR14,
oGR4/oGR10, oGR4/oGR12, oGR4/oGR13, oGR4/oGR14,
oGR5/oGR10, oGR5/oGR12, oGR5/oGR13, oGR5/oGR14,
oGR7/oGR10, oGR7/oGR11, oGR7/oGR12, oGR7/oGR13,
oGR8/oGR10, oGR8/oGR12, oGR8/oGR13, oGR9/oGR10,
oGR9/oGR12, oGR9/oGR13.

c) establishing bona fide sequences.

After initial sequencing of the cloned DNA fragments, more specific primers were designed to isolate the relevant cDNA from the two species.

For the Polythoa species:

oGR21:AAAGGCGTGCCCCTTCCTTTCGCTTTCGA;
oGR22:TGTCAACAGCATTCCAGTATGGCAACAGGGTA;
oGR23:TGAAGAGGGCGTTTGCACCACAAAGAGTG;
oGR24:AAAGGGGAGAAGCTTGACCCCAACGGCC;
oGR25:TTGAAAGCAGTCTGGTTGGCCTTTCTTGA;
oGR26:TGTGGTGCAAACGCCCTCTTCATATTTGAA;
oGR27:CCCTGTTGCCATACTGGAATGCTGTTGAC;
oGR28:AAGGAAGGGGCACGCCTTTAGTGACTGTAAG
oGR29:CTTGCCTTGTCCCTCTCCCGTGATCGTGA;

For the Discosoma species:

oGR39:GGAGAAGGAGAAGGAAAACCATACGAGGG;
oGR40:CCAGTACGGCAACAGGGCATTCACCAAAT;
oGR41:GGGAAAGAACCATGAATTTTGAAGACGGG;
oGR42:CCCCCCATTGGCCCAGTTATGCAGAAGAA;
oGR43:GCCAATGGGGGGAAAGTTCGCACCATCAA;
oGR44:CGCCCCCGTCTTCAAAATTCATGGTTCTT:
oGR45:CCTGTTGCCGTACTGGAACGCTGTTGTCA;
oGR46:TGGGAAGTCTTATGATGGCACCAATACCG;
oGR47:TTCAGGTAACCAAGGGTGGACCTCTGCCA;
oGR48:TGTCAGGCATCCCGAAGACATCGCTGATT:

oGR49:CATGCACTTTGAAGACGGTGGCGTGTGTT;
oGR50:TCATTGGTGATACAACACACGCCACCGTC;
oGR51:CATGACCCTTTCCCATGTAAATCCTTCGGGA;
oGR52:TTGTGGTGACAAAATAGGCCAAGCAAATGGC;
oGR53:GAAATAAAAGGCGACGGTCACGGGAAGCC;
oGR54:CATGGTAACCAAGGGTGGACCCCTGCCAT;
oGR55:AAANCTGTCGTTTCCCGAGGGATTTACAT;
oGR56:TGGCGTGATTTGCAGCNCCAATGATATCA;
oGR57:CGCCACCGNCTTCAAAGTGCATGACCCTT;
oGR58:ANCGGCTATGTCTTCAGGGTGCTTGACAA
oGR59: GGTCCACCCTTGGTTACCATGAGCTTGACGTT.

Following primer combinations are to be envisaged:
oGR21/oGR20, oGR22/oGR20, oGR23/oGR20, oGR24/oGR20,
oGR25/oGR30/OGR31, oGR26/oGR30/OGR31,
oGR27/oGR30/OGR31, oGR28/oGR30/OGR31,
oGR29/oGR30/OGR31, oGR25/oGR16, oGR25/oGR18,
oGR26/oGR16, oGR26/oGR18, oGR27/oGR16, oGR27/oGR18,
oGR28/oGR16, oGR28/oGR18, oGR29/oGR16, oGR29/oGR18,
oGR39/oGR20, oGR40/oGR20, oGR41/oGR20, oGR42/oGR20,
oGR43/oGR30/OGR31, oGR44/oGR30/OGR31,
oGR45/oGR30/OGR31, oGR43/oGR16, oGR43/oGR18,
oGR44/oGR16, oGR44/oGR18, oGR45/oGR16, oGR45/oGR18
oGR46/oGR20, oGR47/oGR20, oGR48/oGR20, oGR49/oGR20,
oGR50/oGR30/OGR31, oGR51/oGR30/OGR31,
oGR52/oGR30/OGR31, oGR50/oGR16, oGR50/oGR18,
oGR51/oGR16, oGR51/oGR18, oGR52/oGR16, oGR52/oGR18,
oGR53/oGR20, oGR54/oGR20, oGR55/oGR20, oGR56/oGR20,
oGR57/oGR30/OGR31, oGR58/oGR30/OGR31,
oGR59/oGR30/OGR31, oGR57/oGR16, oGR57/oGR18,
oGR58/oGR16, oGR58/oGR18, oGR59/oGR16, oGR59/oGR18

d) 3′ and 5′ RACE experiments

To clone the full length cDNA encoding for the fluorescent proteins of the Polythoa species and the Discosoma species, 3′ and 5′ RACE experiments were performed. To facilitate these experiments additional cDNA was prepared. Starting from the RNA isolations as described above, new first strand cDNA synthesis was performed using the SMART PCR cDNA Synthesis Kit (Cat. NO. K1052-1; Clontech). 3′ RACE PCR, was performed according to the manufacturers instructions of the SMART PCR cDNA Synthesis Kit. The 5′ RACE ends of the cDNA fragments were amplified using a step-out RACE strategy (Matz, M. et al. Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res. 27, 1558-1560 (1990)), or according to the manufacturers instructions of the SMART PCR cDNA Synthesis Kit.

The 3′ ends of the Polythoa species were amplified in primary PCR reactions with the primer combinations oGR1-oGR20 and oGR2-oGR20. A sample of the primary PCR reaction was used as a template in nested PCR reactions using primer combinations oGR2-oGR20and oGR3-oGR20 respectively

The 3′ ends of the Discosoma species were amplified in primary PCR reactions with the specific primer combination oGR39/oGR20 after which a nested PCR was performed with primer combinations oGR40/oGR20 or oGR41/oGR20 or oGR42/oGR20. Primary PCR with primers combination oGR41/oGR20 was nested with oGR42/oGR20, and primary PCR reaction with primer combination oGR47/oGR20 was nested with primer combination oGR49/oGR20. Finally PCR reaction with primer combination oGR41/oGR20 was nested with oGR42/oGR20

The primary PCR conditions were: 1′ at 94°C, 30 PCR cycles (30″ at 94°C, 1′ at 55°C and 5′ at 68°C) followed by 5′ at 68°C
The PCR conditions of this nested PCR were as followed: 1′ at 94°C followed by 35 cycles (30″ at 94°C, 1′ at 55°C and 5′ at 72°C) and 5′ at 72°C.

The 5′ ends of the Polythoa species were amplified in primary PCRs with the specific 5′ RACE primers combinations: oGR16/oGR28, oGR16/oGR25, oGR16/oGR26, oGR16/oGR27, oGR16/oGR28 and oGR16/oGR29. The following PCR conditions were used: 1′ at 94°C, 20 PCR cycles (30″ at 94°C, 1′30″ at 72°C (-0.2°C/cycle)), 20 PCR cycles (30″ at 94°C and 1′30″ at 68°C) followed by 5′ at 68°C.

The 5' ends encoding for the Discosoma species fluorescent proteins were amplified according to the Step-Out PCR protocol as mentioned above. Primary PCRs with 5' RACE primers combinations oGR10/oGR30/oGR31 and nested with primers combinations oGR11/oGR30/oGR31 was performed.
Other primary PCR/ nested PCR combinations were:
oGR11/oGR30/oGR31, nested with oGR12/oGR30/oGR31,
oGR12/oGR30/oGR31, nested with oGR13/oGR30/oGR31,
oGR13/oGR30/oGR31, nested with oGR14/oGR30/oGR31,
oGR43/oGR30/oGR31, nested with oGR44/oGR31 or oGR45/oGR31,
oGR44/oGR30/oGR31, nested with oGR45/oGR31,
oGR50/oGR30/oGR31, nested with oGR51/oGR31 or oGR52/oGR31,
oGR51/oGR30/oGR31,nested with oGR52/oGR31,
oGR52/oGR30/oGR31
oGR59/oGR30/oGR31

The primary and nested PCR conditions were: 1' at 94°C, 35 cycles of PCR (30" at 94°C, 1' at 55°C and 2' at 72°C) followed by 5' at 72°C.
The 5' ends of the Discosoma species were also amplified using specific 5' RACE primers combinations oGR43/oGR16, oGR43/oGR18, oGR44/oGR16, oGR44/oGR18, oGR45/oGR16, oGR45/oGR18, oGR50/oGR16, oGR50/oGR18, oGR51/oGR16, oGR51/oGR18, oGR52/oGR16, oGR52/oGR18 and oGR59/oGR16, oGR59/oGR18.
The PCR conditions were an initial denaturation of 1' at 94°C, followed by 20 cycles of touch down PCR (30" at 94°C, 1' at 72°C (-0.2°C/cycle)), followed by 20 cycles of PCR (30" at 94°C and 1' at 68°C) and 5' at 68°C.

All the resulting PCR products of the 3' and 5' RACE PCRs were analyzed on agarose gel and the appropriate DNA bands of interest were isolated and cloned into the pCR-XL-TOPO vector (Cat. NO. K4700-20; Invitrogen) and further analyzed by sequence analysis.

Primers oGR20, oGR16, oGR18, oGR30, oGR31 were provided by the manufacturers and encode for :
oGR20: GTAATACGACTCACTATAGGGCCGCAGTCGACCGTTTTTTTTTTTT

oGR16 AAGCAGTGGTATCAACGCAGAGT
oGR18:AAGCAGTGGTAACAACGCAGAGT
oGR30: GTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
oGR31: GTAATACGACTCACTATAGGGC

e) Cloning of full size cDNA encoding for fluorescent proteins from Anthozoa species.

All cloning experiments were performed using standard protocols as provided by the manufacturers or as described by Ausubel et al. in Current Protocols in Molecular biology, ibid. Isolation of full length cDNA's was also performed using the Titan One tube RT PCR System (Boeringer Mannheim) The reactions were performed according to the manufacturers instructions.

i) Cloning of full size Polythoa 1 GFP cDNA

PCR was performed using specific primer combinations oGR32/oGR34, oGR32/oGR35, oGR33/oGR34 and oGR33/oGR35, and other primer combinations as described above. The resulting fragments were isolated and cloned in appropriate vectors, mainly the pCR-XL-TOPO vector.

The resulting plasmid was designated pGR22 (using primer combination oGR33: CTTGGTGATTTGGGAGAAGGCAGATCGAG and oGR34: CGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG).
Polythoa 1 GFP cDNA was amplified by PCR using plasmid DNA pGR22 as template and the primers: oGR68: CTGGAATTCTATTACTTTGAGTCTACCATCATGAGTGCAATT and oGR72: CGTATCTCGAGCGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG. The resulting PCR products were analyzed by agarose gel electrophoresis and the DNA of interest was isolated and cloned into the pCR-XL-TOPO vector. The resulting plasmid was designated pGR26.

ii) Cloning of full size Polythoa 2 GFP cDNA:

To isolate the full size cDNA clone of the Polythoa species (here designated Polythoa 2), the

Titan One Tube RT-PCR System (Cat. NO. 1888382, Boehringer Mannheim) was used. The reactions were performed according to the manufacturers procedure, using specific primers oGR32 till oGR38. More particularly the following primer combinations were successful:
oGR32/oGR34, oGR32/oGR35, oGR33/oGR34, oGR33/oGR35, oGR36/oGR37 and oGR36/oGR38.

oGR32: ACCTTGGTGATTTGGGAGAAGGCAGATCGAGAG;
oGR33: CTTGGTGATTTGGGAGAAGGCAGATCGAG;
oGR34: CGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG;
oGR35: GAGAAACTTCTTTTTCACTTTGTTGTCGTCTTG;
oGR36: GACACTGGTGATTTGGGAGAAGGCAGATC;
oGR37: ATTGCGAGCCACGGCAACTTCATACAGC;
oGR38: GCCATAATCTGAAGAGGAGAATTGCGAGCCAC).

The resulting PCR products were analyzed by agarose gel electrophoresis and the DNA of interest was isolated and cloned into the pCR-XL-TOPO vector. The resulting plasmids were designated pGR1 (using primers combination oGR32/oGR34) and pGR8 (using primers combination oGR36/oGR38)

iii) Cloning of full size Discosoma 1 GFP cDNA

As in the previous experiments, specific primers were designed based upon the available sequence information resulting from earlier PCR reactions and 3' and 5 ' RACE PCR experiments. The isolation of a full length cDNA is analogous as described above.

iv) Cloning of full size Discosoma 2 GFP cDNA

As in the previous experiments, specific primers were designed based upon the available sequence information resulting from earlier PCR reactions and 3' and 5 ' RACE PCR experiments. The isolation of a full length cDNA is analogous as described above.

2) Cloning of new fluorescent proteins cDNA in expression vectors

a) Cloning of Polytho2 GFP cDNA in prokaryotic expression vector:

Polythoa 2 GFP cDNA was amplified by PCR using plasmid DNA pGR1 as template and the primers: oGR69:CTGGAATTCTCTACCGTCATGAGTGCAATTAAACCAGTCA and oGR70: CGTATCTCGAGATTGCGAGCCACGGCAACTTCATACAGC. or by using plasmid DNA pGR8 as template and the primers oGR68: CTGGAATTCTATTACTTTGAGTCTACCATCATGAGTGCAATT and oGR72: CGTATCTCGAGCGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG.

The PCR product was purified and digested with the restriction enzymes EcoRI and XhoI and cloned in EcoRI/XhoI cloning sites of the expression vector pET32A (Cat. NO. 69015-3; Novagen), the resulting vectors were designated pGR3, and pGR7 respectively. The resulting expression in *E.coli* resulted in visual observation of the fluorescent protein, without induction or UV treatment indicating high expression levels or a fluorescent protein with a high emission amplitude.

b) Cloning of Polytho2 in eukaryotic expression vector:

Polythoa 2 cDNA was amplified by PCR using plasmid DNA pGR8 as template, and the primers combinations oGR69/oGR70 or oGR69/oGR71:
oGR69:CTGGAATTCTCTACCGTCATGAGTGCAATTAAACCAGTCA oGR70:
CGTATCTCGAGATTGCGAGCCACGGCAACTTCATACAGC.
oGR71: CGTATCTCGAGGCCATAATCTGAAGAGGAGAATTGCGAGCCAC
The PCR product was purified and digested with the restriction enzymes EcoRI and XhoI and cloned in EcoRI/XhoI cloning sites of the expression vector pCDNA3 (Invitrogen), the resulting vectors were designated pGR4 and pGR5 respectively.

c) Cloning of Polytho2 in C. elegans expression vector:

Polythoa 2 cDNA was amplified by PCR using plasmid DNA pGR1 as template, and the primers:
oGR74: CGTCGGCGCGCCACCACCATGAGTGCAATTAAGCCAGTTATGAA
and oGR72:
CGTATCTCGAGCGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG.
The PCR product was purified and digested with the restriction enzymes EcoRI and XhoI and cloned in EcoRI/XhoI cloning sites of the expression vector pDW2700, the resulting vector was designated pGR6 .

d) Cloning of Polythoa 1 GFP cDNA in prokaryotic expression vector:
An 752bp EcoRI/XhoI fragment of pGR26 was isolated, purified and ligated into the EcoRI/XhoI cloning sites of the expression vector pCDNA3 (Invitrogen). The resulting vector was designated pGR24. The resulting expression in COSI cells resulted in visual observation of the flurescent protein, after UV treatment.

e) Cloning of Polythoa 1 GFP cDNA in eukaryotic expression vector:
An 752bp EcoRI/XhoI fragment of pGR26 was isolated, purified and ligated into the EcoRI/XhoI cloning sites of the expression vector pET32A (Cat. NO. 69015-3; Novagen). The resulting vector was designated pGR25.

3) Expression of new fluorescent proteins.

a) expression of Polythoa 2 GFP in E. coli

Expression in E.coli was performed according the instructions of the pET32A provider (Novagen). Both the plasmids pGR3 and pGR7 resulted in clear expression in E. coli.

b) expression of Polythoa 2 GFP in Mammalian cells

COS I :African green monkey kidney cell line, standardly cultured in DMEM with Na-pyruvate supplemented with 10% fetal calf serum (Life Technologies) and antibiotics (Pen/Strep; Life Technologies), was transfected with pGR4.
The cells were seeded at a concentration of 1.5×10^4 cells/well in 24-well plate and 7.5×10^4 cells/well in 1 well coverglass and trandsfected the day after with Lipofectamine Plus reagent (GibcoBRL 10964-013), according to the manufacturers instructions.
The following day, the cells where washed twice with PBS (Life Technologies), and complete medium (1ml for 24-well, 3ml for coverglass) was added. Fluorescence of the cells after 24 hours was observed by using UV-light of the microscoop with filter 450-490

(FT510 ;LP520). Both the plasmids pGR4 and pGR5 resulted in clear expression in Cos I cells

c) Expression of Polythoa 2 GFP in C. elegans.

C. elegans wild-type strain was transformed with pGR6 using microinjection techniques known in the art, and described in Methods in Cell Biology, Vol48: C. elegans, Modern biological analysis off an organism, ed. by Epstein and Shakes. pGR6 resulted in clear expression of GFP in *C.elegans*.

4) Mutant fluorescent proteins

To further improve the characteristics of the isolated mutant fluorescent proteins, mutagenesis experiments were performed. Improvements of the fluorescent proteins can be of different nature, such as improved absorption spectra, improved emissions spectra, enhancement of the chromophore, etc.

Site directed mutagenesis can be performed as described in Current protocols in Molecular Biology, ed by Ausubel et al, or as provided in the by the QuickChange Site-Directed Mutagenesis Kit (Stratagene, CA, USA) or by related methods as known in the art. Random mutagenesis, and more particularly molecular evolution techniques can be performed as described by Kunchner and Arnold, 1997, tibtech 15:523-530; Stemmer, 1994, Nature 370:389-391; Stemmer, 1994, Proc. Natl. Acad. Sci. USA 91:10747-10751, or by related methods as known in the art.

During the cloning of the full length cDNA's in the vectors using PCR technology, mutant fluorescent proteins were created. More particularly the plasmids pGR3, pGR4, pGR5, and pGR8 contain a mutant Polythoa 2 N41D GFP, while plasmid pGR7 expresses a Polythoa 2 Q136R GFP mutant and pGR10 is expresses a I106T mutant. The expression experiments described above clearly indicate that mutations introduced in the newly isolated fluorescent proteins, conserves the basic fluorescence property of the protein..

Back mutating towards natural occurring GFP

The mutation Q136R in pGR7 was remutagenised towards the natural occurring Polythoa 2 FP using the QuikChange Site Directed Mutagenesis Kit and the primers
oGR90: GACCCCAACGGCCCAATTATGCAGAAGAAGACCCTGAAATGGGAG and oGR91:
CTCCCATTTCAGGGTCTTCTTCTGCATAATTGGGCCGTTGGGGTC. The resulting vector was designated pGR13

5) Construction of a Polythoa 2-discosoma 1 hybrid GFP

a) Cloning of a Polythoa 2-discosoma 1 hybrid GFP cDNA in prokaryotic expression vector:

The 3′ end of the Discosoma species was amplified in primary PCR reaction with the specific primer combination oGR39/oGR20 as mentioned above (see 1)d). The resulting PCR products were analyzed on agarose gel and the appropriate DNA band of interest was isolated and cloned into the pCR-XL-TOPO vector (Cat.

NO. K4700-20; Invitrogen). The resulting vector was designated pGR17. Plasmid DNA of pGR17 was digested with the restriction enzymes EcoRV and StuI and analyzed on agarose gel. The appropriate band of 525bp was isolated and cloned into the 3736 bp EcoRV fragment of pGR1. The resulting vector was designated pGR14. The resulting expression in E.coli resulted in visual observation of the fluorescent protein, after UV treatment. An 124bp EcoRI-HindIII fragment of pCDNA3.1/hisA (Invitrogen) was isolated, purified and ligated into the 4212bp EcoRI-HindIII fragment of pGR14. The resulting vector was designated pGR15. The resulting expression in E.coli resulted in visual observation of the fluorescent protein, after UV treatment.

b) Cloning of a Polythoa 2-Discosoma 1 hybrid GFP cDNA in eukaryotic expression vector:

Polythoa 2 - Discosoma 1 hybrid cDNA was amplified by PCR using plasmid DNA pGR14 as a template and the primers: oGR69: CTGGAATTCTCTACCGTCATGAGTGCAATTAAACCAGTCA and oGR96: CGTACCTCGAGCCTTTACTTGGTCAGCCGGCTCGGCAGCTTGG. The PCR product was purified and cloned in the cloning vector pCR-XL-TOPO.). The resulting vector was designated pGR19. The 705 bp EcoRI/XhoI fragment of pGR19 was isolated, purified and cloned in EcoRI/XhoI cloning sites of the expression vector pCDNA3 (Invitrogen)). The resulting vector was designated pGR18. The resulting expression in COSI cells resulted in visual observation of the fluorescent protein, after UV treatment.

c) Cloning of a Polythoa 2-discosoma 1 hybrid GFP cDNA in C. elegans expression vector:

Polythoa 2 - Discosoma 1 hybrid cDNA was amplified by PCR using plasmid DNA pGR14 as template, and the primer combination oGR75: CGTCGGCGCGCCATCATGAGTGCAATTAAACCAGTCATGAAGAT and oGR96: CGTACCTCGAGCCTTTACTTGGTCAGCCGGCTCGGCAGCTTGG. The PCR product was purified and cloned in the cloning vector pCR-XL-TOPO. The resulting vector was designated pGR21. The 700 bp AscI/XhoI fragment of pGR21 was isolated, purified and cloned in the AscI/XhoI cloning site of the expression vector pDW2700. The resulting vector was designated pGR20. The resulting expression in C. elegans resulted in visual observation of the fluorescent protein, after UV treatment.

6) Establishing the excitation and emission spectra of the new green fluorescent proteins

Isolation of protein from Polythoa 2 GFP, Polythoa2 N41D GFP and Polythoa 2-discosoma 2 fusion GFP.

The fluorescent proteins were expressed in E. coli from vector pGR3 (N41D), pGR7(Q136R) , pGR13 (back-mutation, natural occurring Polythoa 2 FP), pGR15 (Polythoa-discosoma hybrid protein) and purified using the BugBuster Protein Extraction Reagent (Cat. NO.: 70584-3; Novagen) and the His-Bind Buffer Kit (Cat. NO.:69755-3; Novagen) according to the manufacturers instructions.
The excitation and emission spectra of the samples were then determined. All samples were excited at

490nm. The spectra were corrected for photomultiplierresponse and monochromator transmittance, transformed to wave number and integrated. All experiments were performed in a Amico Bowman Series 2 Luminescence spectrometer (SLM-Amico Spectronic instruments)

1) Synthetic Polythoa 2 Fluorescent protein with optimal codon usage for C. elegans.

To enhance the performance of the fluorescent proteins in organisms other than the Cnidaria species from which these fluorescent proteins were isolated, the codon usage was altered. Although the genetic code is considered to be universal, every organism has its preferred codon usage, which is related to the presence and the expression of tRNA genes, and hence is involved in post-transcriptional expression regulation. Such optimal codon usage has been determined for many organisms, including *E.coli* (Dong et al., 1996, J. Mol. Biol. 260:649-663), B. subtilis (Kanaya et al., 1999, Gene 238:143-155), Drosophila (Moriyama et al., 1997, J. Mol. Evol. 45:514-523) Saccharomyces (Percudani et al., 1997, J. Mol Biol.268:322-330), C. elegans (Stenico, et al., 1994, NAR 22:2437-2446). An overview of codon usage in these and other organisms can be found in Duret et al., 1999, Proc. Natl. Acad. Sci. U.S.A. 96: 4482-4487 and in Ikemura, 1985, Mol. Biol. Evol. 2:13-43.

The synthetic 922 bp gene was amplified using herculase-polymerase at Entechelon, Germany and was delivered as a ligation product. This product was cloned into pCR-XL-TOPO (pGR16). The 888bp FseI-NheI

fragment of pGR16 was cloned into the FseI/NheI cloning sites of the expression vector pDW2721 and the resulting vector was designated pGR10. This plasmid was injected in C. elegans, and clearly resulted in fluorescence

2) Synthetic introns in worm construct

In many organisms, such as in C.elegans, the introduction of synthetic introns results in enhancements of expression levels (Fire et al. , 1990, Gene 93:189-98, end references therein).
An example is hereby included of a Polythoa 2 fluorescent protein improved for optimal codon usage for C. elegans and with synthetic C.elegans introns. Such synthetic genes can be made easily by a person skilled in the art, or be ordered by companies such as Entelechon, Rgensburg, Germany.

Fusion proteins

GFP proteins have been used for many purposes in biological research. The main use nevertheless has been the expression pattern of proteins in cells and multi-cellular organisms, and the subcellular localization or trafficking of proteins. To determine the expression pattern of a protein using GFP's it suffices in principle to make a fusion between the promoter of the gene of interest and the GFP. Upon introducing a vector with this promoter GFP fusion into the studied cell or organism, the expression induced by the promoter can easily be monitored by following the GFP expression. To monitor the subcellular expression of a protein, it suffices to

make a fusion between the protein of interest and the GFP protein. this can be done at the N-terminal site or at the C-terminal site of the GFP protein, and even internal fusions are possible . Plasmids pGR3, pGR7 and pGR13 are good examples of such fusion proteins as they contain a 109 throredoxin Aminoacid fragment in fusion with the Polythoa 2 GFP. This fusion protein shows clear fluorescence.

TABLE 1

primers	5'---3'
oGR1	CACCACATGGAAGGAWRYKTNRAYGG
oGR2	ACCACATGGAAGGATGCKTNRAYGGNCA
oGR3	AATTTGTGATCAAGGGCRARGGNRWNGG
oGR4	GTGATCAAAGGTGGACCNYTNCCNTT
oGR5	GACATATTGTCAACAGAGTTYMANTAYG
oGR6	CATATTGTCAACAGAGTTYMANTAYGG
oGR7	ATCCTGACGACATACCAGAYTAYHWNAA
oGR8	GACTATTTCAAGCAGTCGTKYCCNGMNGG
oGR9	CATGGGAAAGGTCCTTGCAYTWYGARGA
oGR10	GGTGACATCTCCTTTCARNAYNCC
oGR11	CATATTCTCAGTGGANGSNTCCCA
oGR12	CACAGGTCCATCGSNAGGRAARTT
oGR13	CCATCGGCAGGAAARTTNANNCC
oGR14	TGAATACCCTGTTTCCRTANTKRAA
oGR16	AAGCAGTGGTATCAACGCAGAGT
oGR18	AAGCAGTGGTAACAACGCAGAGT
oGR20	GTAATACGACTCACTATAGGGCCGCAGTCGACCGTTTTTTTTTTTTT
oGR21	AAAGGCGTGCCCCTTCCTTTCGCTTTCGA
oGR22	TGTCAACAGCATTCCAGTATGGCAACAGGGTA
oGR23	TGAAGAGGGCGTTTGCACCACAAAGAGTG
oGR24	AAAGGGGAGAAGCTTGACCCCAACGGCC
oGR25	TTGAAAGCAGTCTGGTTGGCCTTTCTTGA
oGR26	TGTGGTGCAAACGCCCTCTTCATATTTGAA
oGR27	CCCTGTTGCCATACTGGAATGCTGTTGAC
oGR28	AAGGAAGGGGCACGCCTTTAGTGACTGTAAG
oGR29	CTTGCCTTGTCCCTCTCCCGTGATCGTGA
oGR30	GTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
oGR31	GTAATACGACTCACTATAGGGC
oGR32	ACCTTGGTGATTTGGGAGAAGGCAGATCGAGAG
oGR33	CTTGGTGATTTGGGAGAAGGCAGATCGAG
oGR34	CGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG
oGR35	GAGAAACTTCTTTTTCACTTTGTTGTCGTCTTG
oGR36	GACACTGGTGATTTGGGAGAAGGCAGATC
oGR37	ATTGCGAGCCACGGCAACTTCATACAGC
oGR38	GCCATAATCTGAAGAGGAGAATTGCGAGCCAC
oGR39	GGAGAAGGAGAAGGAAAACCATACGAGGG
oGR40	CCAGTACGGCAACAGGGCATTCACCAAAT
oGR41	GGGAAAGAACCATGAATTTTGAAGACGGG
oGR42	CCCCCCATTGGCCCAGTTATGCAGAAGAA
oGR43	GCCAATGGGGGGAAAGTTCGCACCATCAA
oGR44	CGCCCCCGTCTTCAAAATTCATGGTTCTT
oGR45	CCTGTTGCCGTACTGGAACGCTGTTGTCA
oGR46	TGGGAAGTCTTATGATGGCACCAATACCG
oGR47	TTCAGGTAACCAAGGGTGGACCTCTGCCA
oGR48	TGTCAGGCATCCCGAAGACATCGCTGATT
oGR49	CATGCACTTTGAAGACGGTGGCGTGTGTT

oGR50	TCATTGGTGATACAACACACGCCACCGTC
oGR51	CATGACCCTTTCCCATGTAAATCCTTCGGGA
oGR52	TTGTGGTGACAAAATAGGCCAAGCAAATGGC
oGR53	GAAATAAAAGGCGACGGTCACGGGAAGCC
oGR54	CATGGTAACCAAGGGTGGACCCCTGCCAT
oGR55	AAANCTGTCGTTTCCCGAGGGATTTACAT
oGR56	TGGCGTGATTTGCAGCNCCAATGATATCA
oGR57	CGCCACCGNCTTCAAAGTGCATGACCCTT
oGR58	ANCGGCTATGTCTTCAGGGTGCTTGACAA
oGR59	GGTCCACCCTTGGTTACCATGAGCTTGACGTT
oGR68	CTGGAATTCTATTACTTTGAGTCTACCATCATGAGTGCAATT
oGR69	CTGGAATTCTCTACCGTCATGAGTGCAATTAAACCAGTCA
oGR70	CGTATCTCGAGATTGCGAGCCACGGCAACTTCATACAGC
oGR71	CGTATCTCGAGGCCATAATCTGAAGAGGAGAATTGCGAGCCAC
oGR72	CGTATCTCGAGCGTCTTGGCTTTTCGTTAAGCCTTTACTGGGG
oGR74	CGTCGGCGCGCCACCACCATGAGTGCAATTAAGCCAGTTATGAA
oGR75	CGTCGGCGCGCCATCATGAGTGCAATTAAACCAGTCATGAAGAT
oGR90	GACCCCAACGGCCCAATTATGCAGAAGAAGACCCTGAAATGGGAG
oGR91	CTCCCATTTCAGGGTCTTCTTCTGCATAATTGGGCCGTTGGGGTC
oGR96	CGTACCTCGAGCCTTTACTTGGTCAGCCGGCTCGGCAGCTTGG
oGR97	CGTACCTCGAGGATGGATCCTTTACTTGGTCAGCCG

Tabe 2 Primer combinations

oGR1	oGR10	
oGR1	oGR11	
oGR1	oGR12	
oGR1	oGR13	
oGR1	oGR14	
oGR1	oGR20	
oGR10	oGR30	oGR31
oGR11	oGR30	oGR31
oGR12	oGR30	oGR31
oGR13	oGR30	oGR31
oGR14	oGR30	oGR31
oGR16	oGR25	
oGR16	oGR26	
oGR16	oGR27	
oGR16	oGR28	
oGR16	oGR29	
oGR2	oGR10	
oGR2	oGR11	
oGR2	oGR12	
oGR2	oGR13	
oGR2	oGR14	
oGR2	oGR20	
oGR21	oGR20	
oGR22	oGR20	
oGR23	oGR20	
oGR24	oGR20	
oGR25	oGR16	
oGR25	oGR18	
oGR25	oGR30	oGR31
oGR26	oGR16	
oGR26	oGR18	
oGR26	oGR30	oGR31
oGR27	oGR16	
oGR27	oGR18	
oGR27	oGR30	oGR31
oGR28	oGR16	
oGR28	oGR18	
oGR28	oGR30	oGR31
oGR29	oGR16	
oGR29	oGR18	
oGR29	oGR30	oGR31
oGR3	oGR10	
oGR3	oGR11	
oGR3	oGR12	

oGR3	oGR13	
oGR3	oGR14	
oGR3	oGR20	
oGR32	oGR33	
oGR32	oGR34	
oGR32	oGR35	
oGR33	oGR34	
oGR33	oGR35	
oGR34	oGR35	
oGR36	oGR37	
oGR36	oGR38	
oGR39	oGR20	
oGR4	oGR10	
oGR4	oGR11	
oGR4	oGR12	
oGR4	oGR13	
oGR4	oGR14	
oGR40	oGR20	
oGR41	oGR20	
oGR42	oGR20	
oGR43	oGR16	
oGR43	oGR18	
oGR43	oGR30	oGR31
oGR44	oGR16	
oGR44	oGR18	
oGR44	oGR30	oGR31
oGR44	oGR31	
oGR45	oGR16	
oGR45	oGR18	
oGR45	oGR30	oGR31
oGR45	oGR31	
oGR46	oGR20	
oGR47	oGR20	
oGR48	oGR20	
oGR49	oGR20	
oGR5	oGR10	
oGR5	oGR11	
oGR5	oGR12	
oGR5	oGR13	
oGR5	oGR14	
oGR50	oGR16	
oGR50	oGR18	
oGR50	oGR30	oGR31
oGR51	oGR16	

oGR51	oGR18	
oGR51	oGR30	oGR31
oGR51	oGR31	
oGR52	oGR16	
oGR52	oGR18	
oGR52	oGR30	oGR31
oGR52	oGR31	
oGR53	oGR20	
oGR54	oGR20	
oGR55	oGR20	
oGR56	oGR20	
oGR57	oGR16	
oGR57	oGR18	
oGR57	oGR30	oGR31
oGR58	oGR16	
oGR58	oGR18	
oGR58	oGR30	oGR31
oGR59	oGR16	
oGR59	oGR18	
oGR59	oGR30	oGR31
oGR6	oGR10	
oGR6	oGR11	
oGR6	oGR12	
oGR6	oGR13	
oGR68	oGR72	
oGR69	oGR70	
oGR69	oGR71	
oGR69	oGR96	
oGR69	oGR97	
oGR7	oGR10	
oGR7	oGR11	
oGR7	oGR12	
oGR7	oGR13	
oGR72	oGR74	
oGR75	oGR96	
oGR75	oGR97	
oGR8	oGR10	
oGR8	oGR11	
oGR8	oGR12	
oGR8	oGR13	
oGR9	oGR10	
oGR9	oGR11	
oGR9	oGR12	
oGR9	oGR13	

CLAIMS:

1. An isolated nucleic acid molecule encoding a fluorescent protein comprising an amino acid sequence illustrated in any of the polypeptide sequences of figures, 3(a) to 3(d) or functional equivalents, fragments or variants thereof.

2. An isolated nucleic acid molecule encoding a protein capable of emitting fluorescence upon irradiation by incident light, wherein said maximal absorbance of said incident light is in the range 440-480 nm, and maximal fluorescence emission is in the range 470-510 nm.

3. An isolated nucleic acid molecule according to claim 2, wherein said molecule encodes a protein having an amino acid sequence as depicted in any of the polypeptide sequences of Figures 3(a) to 3(d).

4. An isolated nucleic acid molecule according to claim 1 wherein said fluorescent protein comprises an amino acid sequence having combined polypeptide sequences from at least 2 of the polypeptide sequences depicted in Figures 3(a) to 3(d).

5. An isolated nucleic acid molecule according to claim 4 wherein said protein comprises a Polythoa 2-Discosoma 1 hybrid having the sequence illustrated in Figure 7.

6. An isolated nucleic acid molecule encoding a fusion protein comprising an amino acid sequence

depicted in any of Figures 3(a) to 3(d) together with a nucleotide sequence encoding a protein of interest.

7. An isolated nucleic acid molecule according to claim 6 wherein said fusion protein comprises the amino acid sequences depicted in Figures 4 and 5.

8. An isolated nucleic acid molecule according to claim 5 wherein said protein of interest is an antibody.

9. An isolated nucleic acid molecule according to any of claims 1 to 8, which is a DNA molecule.

10. An isolated nucleic acid molecule according to claim 9, wherein said DNA molecule is cDNA.

11. An isolated nucleic acid molecule according to any of claims 1 to 10, which is derived from an Anthozoa species.

12. An isolated nucleic acid molecule according to claim 11, wherein said Anthozoa species is any of a Polythoa or Discosoma species.

13. An isolated nucleic acid molecule according to any preceding claim, wherein said molecule comprises a nucleotide sequence which has at least 70, preferably at least 80, more preferably at least 90 and even more preferably at least 95% sequence identity to the nucleic acid sequences depicted in Figure 1.

14. An isolated nucleic acid molecule according to any preceding claim, wherein said nucleic acid molecule comprises any of the nucleic acid sequences depicted in Figure 1.

15. An isolated nucleic acid molecule according to claim 13 comprising any of the nucleotide sequences depicted in Figure 2(a) or 2(b).

16. An antisense molecule capable of hybridising to a nucleic acid molecule according to any of claims 1 to 13, under conditions of high stringency.

17. An isolated fluorescent protein or functional equivalent, derivative or variant thereof encoded by a nucleic acid molecule according to any of claims 1 to 13.

18. An isolated fluorescent protein capable of emitting fluorescence upon irradiation by incident light wherein the maximal absorbance of said incident light is in the range 440-480 nm, and maximal fluorescence emission is in the range 470-510 nm.

19. An isolated fluorescent protein comprising an amino acid sequence which has at least 70, preferably at least 80, more preferably at least 90 and even more preferably at least 95% sequence identifying to the amino acid sequence depicted in Figures 3 to 8.

20. An isolated fluorescent protein comprising an amino acid sequence corresponding substantially the

polypeptide sequences depicted in any of Figures 3 to 8.

21. An isolated fusion fluorescent protein comprising a fluorescent protein according to any of claims 16 to 20 together with the amino acid sequence of a protein or polypeptide of interest.

22. A fluorescently labelled antibody or a paratope thereof coupled to a fluorescent protein according to any of claims 16 to 20.

23. An expression vector comprising any of the nucleic acid molecules according to claims 1 to 15.

24. An expression vector comprising any of the plasmid sequences depicted in Figures 9 to 14.

25. An expression vector comprising the sequences of any of plasmids pGR8 to pGR20.

26. A host cell transformed or transfected with an expression vector according to any of claims 23 to 25.

27. A prokaryotic cell transformed or transfected with any of expression vectors pGR3, pGR7 depicted in Figures 9 and 13 or pGR13.

28. A prokaryotic cell according to claim 25 which is *E.coli*.

29. A eukaryotic cell transformed or transfected with an expression vector corresponding

substantially to the plasmids designated pGR4 or PGR5 in Figures 10 or 11.

30. A transgenic cell tissue or non-human organism comprising a transgene capable of expressing a fluorescent protein according to any of claims 17 to 21 or an antibody according to claim 22.

31. A transgenic cell, tissue or non-human organism according to claim 30, wherein said transgene is included in an expression vector.

32. A transgenic cell, tissue or non-human organism according to claim 31, wherein said vector is one according to claim 23.

33. A transgenic cell, tissue or non-human organism wherein said non-human organism is C-elegans and said transgene substantially corresponds to a nucleotide sequence as depicted in Figure 12.

34. A fluorescent protein, or a functional equivalent, derivative or bioprecursor thereof, expressed by a cell, tissue or organism according to any of claims 27 to 33.

35. A process for producing the protein of any one of claims 17 to 21, comprising the steps of cultivating a cell tissue or organism according to any of claims 24 to 33 under conditions suitable for expression of the protein and optionally recovering the expressed protein.

36. An oligonucleotide probe comprising at least about 10 nucleotides of a nucleotide sequence that is capable of selectively hybridising to a nucleic acid molecule according to any of claims 1 to 15.

37. A method for selecting cells capable of expressing a protein of interest, comprising introducing into said cells a vector comprising the nucleotide sequence of a fluorescent protein according to any of claims 17 to 22 operatively linked to a promoter or regulatory region of the protein of interest, cultivating the cell under conditions necessary for expressing the protein of interest and monitoring for any fluorescence following expression of said fluorescent protein.

38. A method for producing fluorescence resonance energy transfer comprising;

providing a donor molecule comprising a fluorescent protein according to any of claims 17 to 21;

providing an appropriate acceptor molecule for the fluorescent protein; and

bringing the donor molecule and acceptor molecule into sufficiently close contact to allow fluorescent resonance energy transfer.

39. A method for producing fluorescence resonance energy transfer comprising;

providing an acceptor molecule comprising a fluorescent protein according to any of claims 17 to 21;

providing an appropriate donor molecule for the fluorescent protein; and

bringing the donor molecule and acceptor molecule into sufficiently close contact to allow fluorescence resonance energy transfer.

40. A microscopic nematode comprising a transgene capable of expressing a fluorescent protein according to any of claims 17 to 20.

41. A nematode according to claim 40 which is *C.elgans*.

42. A fluorescent protein obtainable from the coral species Anthozoa.

43. A fluorescent protein according to claim 41 which is obtainable from Discosoma or Polythoa.

44. A fluorescent protein according to claim 42 or 43 which is capable of emitting fluorescence upon irradiation by incident light wherein the maximal absorbance of said incident light is in the range 440-480 nm, and maximal fluorescence emission is in the range 470-510 nm.

45. A fluorescent protein according to claim 42 or 43 comprising an amino acid sequence which has at least 70, preferably at least 80, more preferably at least 90 and even more preferably at least 95% sequence identifying to the amino acid sequence depicted in Figures 3 to 8

CDNA fragment of polythoa 1 encoding for a fluorescent protein
Start codon ATG and stop codon TAA are indicated.

```
   1 acgcggggat tcaccttggt gatttgggag aaggcagatc gagagcaaga gtcagtgtaa
  61 taacttactt tgagtctacc atcatgagtg caattaagcc agttatgaag gtagaattgg
 121 tcatggaagg aaatgtgaac gggcacaagt tcacgattac aggagaggga caaggcaagc
 181 cttacgaggg aactcacact ctaaacctta cagtcacaaa aggcgggccc cttcctttcg
 241 cttacgatat cttgtcarca gcattccagt acggcaacag ggtatttacc aaatacccag
 301 aagatatacc ggactatttc aagcagacct ttccagaagg atattcgtgg gaaagaactt
 361 tcaaatatga cgagggcctt tgcaccacaa aaagtgacat atgcctcaag aaaggcgaac
 421 cggactgctt tcaatacaaa atttactttg aagggaagaa ccttggcccc agcggtccaa
 481 ttatgcagaa gaagaccctg aaatgggagc catccactga gaggatgtac atggacgtgg
 541 ataaagacgg tgcaaaggtg ctgaagggcg atgataatgc ggccctgttg cttgaaggag
 601 gtggccatta tcgttgtgac ttcaatagta tttacaaggc gaagaaaact gggtccttgc
 661 cagcatatca ctggatagac caccgcattg agattttgag ccacgataaa gattacaaca
 721 aggttacaat gcatgaattt gccgctgctc gtaattctcc ttttccgata atggcgcccc
 781 agtaaaggct taacgaaaag ccaagacgac aacaaagtga aaaagaagtt tctcgtttac
 841 ttttttctga aggcatttat cactaattag cttttgatag ttttgattca cggattcgat
 901 ccatgaattt cttagggact agctctagaa taaatgattg tgaaacaaaa actagttttc
 961 atattttgcg agatttttca cttcataaag acagactttt taaactcagt tgtagccaaa
1021 tacaaataag gaaagtgtat taagaattaa acaaacttgt tgtggaaaaa taataaaaac
1081 ggtcgactgc ggccctataa tgagtcgtat tac
```

(a)

CDNA fragment of polythoa 2 encoding for a fluorescent protein
Start codon ATG and stop codon TAA are indicated.

```
   1 acgcggggac actggtgatt tgggagaagg cagatcgaga gcaagagtca gtgtaataac
  61 ttactttgag tctaccgtca tgagtgcaat taaaccagtc atgaagattg aattggtcat
 121 ggaaggagag gtgaacgggc acaagttcac gatcacggga gagggacaag gcaagcctta
 181 cgagggaaca cagactctaa accttacagt cactaaaggc gtgccccttc ctttcgcttt
 241 cgatatcttg tcaacagcat tccagtatgg caacagggta tttaccaaat acccagatga
 301 tataccggac tatttcaagc agacctttcc ggaaggatat tcgtgggaaa gaactttcaa
 361 atatgaagag ggcgtttgca ccacaaagag tgacataagc ctcaagaaag gccaaccaga
 421 ctgctttcaa tataaaatta actttaaagg ggagaagctt gaccccaacg gcccaattat
 481 gcagaagaag accctgaaat gggagccatc cactgagagg atgtacatgg acgtggataa
 541 agacggtgca aaggtgctga agggcgatgt taatgcggcc ctgttgcttg aaggaggtgg
 601 ccattatcgt tgtgacttta acagtactta caaggcgaag aaaactgtgt ccttcccagc
 661 atatcacttt gtggaccacc gcattgagat tttgagccac aatacggatt acagcaaggt
 721 tacactgtat gaagttgccg tggctcgcaa ttctcctctt cagattatgg cgccccagta
 781 aaaggcttaac gaaacgccaa tacgacaaca aagtgaaaaa caagttttc gttatttttt
 841 tctgaaagca tttatcacta attagctttt gatagttttg attcacggat tcgatccgga
 901 atttaatagg gactagctct agtctagaat aaacgattgt gtaacaaaaa ctagctttca
 961 taatttgcgg gatttttcac ttcataaaga cagacttttt aaactcagtt gtagccaaat
1021 acaaataagg aaagcgtatt aagaattaaa caaacttgtt gtcgaaaaaa aaaaaaacgg
1081 tcgattgcgg ccctatagtg agtcgtatta c
```

(b)

CDNA fragment of discosoma 1 encoding for a fluorescent protein
stop codon TAA are indicated.

```
   1 caccacatgg aaggaagtgt ggacgggcaa aatttcgtga tcactggaga aggagaagga
  61 aaaccatacg agggaacaca tgttatagac ctgcaagtcg ttgaaggcgg acctctgcgt
 121 ttcgcttacg atatcttgac aacagcgttc cagtacggca acagggcatt caccaaatac
 181 ccatcagata ttcctgacta tttcaagcag acttttcctc aagggtatac atgggaaaga
 241 accatgcact ttgaagacgg tggcgtgtgt accgtcaata gcgacgtaag cctgaaaagc
 301 ggctgttttg agtataaaat tcgttttgat ggtgagaact ttccccccaa tggcccagtt
 361 atgcagaaga agactgtgaa atgggagcca tccactgaga acatgtatga gcatgatggg
 421 atgctgaagg gtgatgttag cagaactctg ttgcttgaag gaggtggcca ttaccaatgc
 481 gactttaaaa ctatttacaa agcgaaggac agccagggaa tcaagatgcc agaatatcac
 541 tttgtggacc accgcattga gattttgagc catgacaaag attacaagat ggtcaaggtg
```

(c)

Fig 1

```
601 tatgagattg ccgaagctca ctattccaag ctgccgagcc ggctgaccaa gtaaaggcct
661 aaggaaaagc caacaagcca acaaggagga aaaaatacta gtgtttctag tacagttttt
721 taagccattt actaggaatt agtttttaat acttcagatc gtttcgggat ttgttagaga
781 ttagcttacg aaaactgata ctcctagagt ttctagtatt gtttttaagc catttactcg
841 gaattagttt ttgatacttt agatcgtttc ggaatttgtt agagtttagc tttaaaaaaa
901 tactagactg
```

(c) (contd)

CDNA fragment of discosoma 2 encoding for a fluorescent protein

```
  1 caccacatgg aaggaagtgt tgacggccac tactttgaaa ttaaaggcaa tggatatggg
 61 aagtcttatg atggcaccaa taccgtaaag cttcaggtaa ccaagggtgg acctctgcca
121 tttgcttggc ctattttgtc accacaattt caatatggaa acaagatatt tgtcaggcat
181 cccgaagaca tcgctgatta taaaaagctg tcatttcccg aaggatttac atgggaaagg
241 gtcatgcact ttgaagacgg tggcgtgtgt tgtatcacca atgatatcag tttggaaggc
301 aactgtttca tctaccacat caatttcatt ggcttgaact ttccttccga tggacctgtg
```

(d)

Fig 1 (cont'd)

DNA fragment encoding polythoa 2 with optimal codon usage for C. elegans as in plasmid pGR10.

(a)

```
   1 atgtccgcta tcaagccagt catgaagatc gagctcgtca tggagggaga ggtcaacgga
  61 cacaagttca ccatcaccgg agagggacag ggaaagccat acgagggaac ccagaccctc
 121 aacctcaccg tcaccaaggg agtcccactc ccattcgctt tcgacatcct ctccaccgct
 181 ttccagtacg gaaaccgtgt cttcaccaag tacccagacg acatcccaga ctacttcaag
 241 cagaccttcc cagagggata ctcctgggag cgtaccttca agtacgagga gggagtctgc
 301 accaccaagt ccgacatctc cctcaagaag ggacagccag actgcttcca gtacaagatc
 361 aacttcaagg gagagaagct cgacccaaac ggaccaatca tgcagaagaa gaccctcaag
 421 tgggagccat ccaccgagcg tatgtacatg gacgtcgaca aggacggagc taaggtcctc
 481 aagggagacg tcaacgctgc tctcctcctc gagggaggag gacactaccg ttgcgacttc
 541 aactccacct acaaggctaa gaagaccgtc tccttcccag cttaccactt cgtcgaccac
 601 cgtatcgaga tcctctccca caacaccgac tactccaagg tcaccctcta cgaggtcgct
 661 gtcgctcgta actccccact ccagatcatg gctccacag
```

DNA fragment encoding polythoa 2 with optimal codon usage for C.elegans further including introns. the introns are underlined. Furthermore the starting codon ATG is preceded by a 5' UTR containing an Kozak site.

(b)

```
   1 tggctagcgt cgacggtacc ggtagaaaaa atgtccgcta tcaagccagt catgaagatc
  61 gagctcgtca tggagggaga ggtcaacgga cacaagttca ccatcaccgg agagggacag
 121 ggaaagccat acgagggaac ccagaccctc aacctcaccg tcaccaaggg agtcccactc
 181 ccattcgctt tcgtaagttt aaacatatat atactaacta accctgatta tttaaatttt
 241 caggacatcc tctccaccgc tttccagtac ggaaaccgtg tcttcaccaa gtacccagac
 301 gacatcccag actacttcaa gcagaccttc ccagagggat actcctggga gcgtaccttc
 361 aagtacgagg agggagtctg caccaccaag taagtttaaa cagttcggta ctaactaacc
 421 atacatattt aaattttcag gtccgacatc tccctcaaga agggacagcc agactgcttc
 481 cagtacaaga tcaacttcaa gggagagaag ctcgacccaa acggaccaat catgcagaag
 541 aagaccctca agtgggagcc atccaccgag cgtatgtaca tggacgtcga caaggacgga
 601 gctaaggtcc tcaaggtaag tttaaacttg gacttactaa ctaacggatt atatttaaat
 661 tttcagggag acgtcaacgc tgctctcctc ctcgagggag gaggacacta ccgttgcgac
 721 ttcaactcca cctacaaggc taagaagacc gtctccttcc cagcttacca cttcgtcgac
 781 caccgtatcg agatcctctc ccacaacacc gactactcca aggtcaccct ctacgaggtc
 841 gctgtcgctc gtaactcccc actccagatc atggctccac agtagggccg gccgagctcc
 901 gcatcggccg ctgtc
```

Fig 2

Protein sequence of polythoa 1

```
  1 MSAIKPVMKV ELVMEGNVNG HKFTITGEGQ GKPYEGTHTL NLTVTKGGPL PFAYDILSXA
 61 FQYGNRVFTK YPEDIPDYFK QTFPEGYSWE RTFKYDEGLC TTKSDICLKK GEPDCFQYKI
121 YFEGKNLGPS GPIMQKKTLK WEPSTERMYM DVDKDGAKVL KGDDNAALLL EGGGHYRCDF
181 NSIYKAKKTG SLPAYHWIDH RIEILSHDKD YNKVTMHEFA AARNSPFPIM APQ*
```

a

Protein sequence of polythoa 2

```
  1 MSAIKPVMKI ELVMEGEVNG HKFTITGEGQ GKPYEGTQTL NLTVTKGVPL PFAFDILSTA
 61 FQYGNRVFTK YPDDIPDYFK QTFPEGYSWE RTFKYEEGVC TTKSDISLKK GQPDCFQYKI
121 NFKGEKLDPN GPIMQKKTLK WEPSTERMYM DVDKDGAKVL KGDVNAALLL EGGGHYRCDF
181 NSTYKAKKTV SFPAYHFVDH RIEILSHNTD YSKVTLYEVA VARNSPLQIM APQ*
```

b

Protein sequence of the N-terminal part of discosoma 1

```
  1 HHMEGSVDGQ NFVITGEGEG KPYEGTHVID LQVVEGGPLR FAYDILTTAF QYGNRAFTKY
 61 PSDIPDYFKQ TFPQGYTWER TMHFEDGGVC TVNSDVSLKS GCFEYKIRFD GENFPPNGPV
121 MQKKTVKWEP STENMYEHDG MLKGDVSRTL LLEGGGHYQC DFKTIYKAKD SQGIKMPEYH
181 FVDHRIEILS HDKDYKMVKV YEIAEAHYSK LPSRLTK*
```

c

Protein sequence of an internal part of discosoma 2

```
  1 HHMEGSVDGH YFEIKGNGYG KSYDGTNTVK LQVTKGGPLP FAWPILSPQF QYGNKIFVRH
 61 PEDIADYKKL SFPEGFTWER VMHFEDGGVC CITNDISLEG NCFIYHINFI GLNFPSDGPV
```

d

Fig 3

polythoa 2 fluorescent fusion protein in pGR3

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPK
YGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAK
FERQHMDSPDLGTDDDDKAMADIGSEFSTVMSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTL
DLTVTKGVPLPFAFDILSTAFQYGNRVFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKK
GQPDCFQYKINFKGEKLDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDF
NSTYKAKKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNLEHHHHHH*

Fig 4

Polythoa 2 fluorescent fusion protein in pGR7

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPK
YGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAK
FERQHMDSPDLGTDDDDKAMADIGSEFYYFESTIMSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEG
TQTLNLTVTKGVPLPFAFDILSTAFQYGNRVFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDI
SLKKGQPDCFQYKINFKGEKLDPNGPIMRKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHY
RCDFNSTYKAKKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ*

Fig 5

```
(a) pGR1        ..........................................................
(b) pGR10       ..........................................................
(c) pGR13       MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLN
(d) pGR16       ..........................................................
(e) pGR3        MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLN
(f) pGR4        ..........................................................
(g) pGR5        ..........................................................
(h) pGR6        ..........................................................
(i) pGR7        MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLN
(j) pGR8        ..........................................................
(k) POLYTHOA2   ..........................................................
    consensus

pGR1        ..........................................................
pGR10       ..........................................................
pGR13       IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHH
pGR16       ..........................................................
pGR3        IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHH
pGR4        ..........................................................
pGR5        ..........................................................
pGR6        ..........................................................
pGR7        IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHH
pGR8        ..........................................................
POLYTHOA2   ..........................................................
consensus

pGR1        ..........................................................MSAIKP
pGR10       ..........................................................MSAIKP
pGR13       HHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSEFYYFESTIMSAIKP
pGR16       ..........................................................MSAIKP
pGR3        HHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSEFS....TVMSAIKP
pGR4        ..........................................................MSAIKP
pGR5        ..........................................................MSAIKP
pGR6        ..........................................................MSAIKP
pGR7        HHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSEFYYFESTIMSAIKP
pGR8        ..........................................................MSAIKP
POLYTHOA2   ..........................................................MSAIKP
consensus                                                             MSAIKP

pGR1        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
pGR10       VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
pGR13       VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
pGR16       VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
pGR3        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLDLTVTKGVPLPFAFDILSTAFQYGNR
pGR4        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLDLTVTKGVPLPFAFDILSTAFQYGNR
pGR5        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLDLTVTKGVPLPFAFDILSTAFQYGNR
pGR6        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
pGR7        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
pGR8        VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLDLTVTKGVPLPFAFDILSTAFQYGNR
POLYTHOA2   VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVTKGVPLPFAFDILSTAFQYGNR
consensus   VMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLnLTVTKGVPLPFAFDILSTAFQYGNR

pGR1        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR10       VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDTSLKKGQPDCFQYKINFKGEK
pGR13       VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR16       VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDTSLKKGQPDCFQYKINFKGEK
pGR3        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR4        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR5        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR6        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR7        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
pGR8        VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
POLYTHOA2   VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDISLKKGQPDCFQYKINFKGEK
consensus   VFTKYPDDIPDYFKQTFPEGYSWERTFKYEEGVCTTKSDiSLKKGQPDCFQYKINFKGEK
```

Fig 6

```
pGR1        LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR10       LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR13       LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR16       LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR3        LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR4        LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR5        LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR6        LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR7        LDPNGPIMRKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
pGR8        LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
POLYTHOA2   LDPNGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA
consensus   LDPNGPIMqKKTLKWEPSTERMYMDVDKDGAKVLKGDVNAALLLEGGGHYRCDFNSTYKA

pGR1        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
pGR10       KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
pGR13       KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
pGR16       KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
pGR3        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNLEHHHHHH..............
pGR4        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNLEHASRGPYSIVSPKC......
pGR5        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIIASSMHLEGPIL.....
pGR6        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
pGR7        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
pGR8        KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPFQIMARANSSTLAAVTSGSE-
POLYTHOA2   KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNSPLQIMAPQ.............
consensus   KKTVSFPAYHFVDHRIEILSHNTDYSKVTLYEVAVARNsplqimapq

pGR1        .......
pGR10       .......
pGR13       .......
pGR16       .......
pGR3        .......
pGR4        .......
pGR5        .......
pGR6        .......
pGR7        .......
pGR8        LGTKLDA
POLYTHOA2   .......
consensus
```

Fig 6 (cont'd)

```
hybridPolyth2-Disco1 ..........................................................
pGR14                ..........................................................
pGR15                MKDDIKKLTMGGSHHHHHHGMASMTGGQQMGRDLYDDDDKVPRIQCGGIRPYLGDLGEGR
pGR17                ..........................................................
pGR18                ..........................................................
pGR19                ..........................................................
pGR20                ..........................................................
pGR21                ..........................................................
consensus

hybridPolyth2-Disco1 ..............MSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
pGR14                ..............MSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
pGR15                SRARVSVITYFESTVMSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
pGR17                ..........................................GEGEGKPYEGTHVVDLQVV
pGR18                ..............MSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
pGR19                ..............MSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
pGR20                ..............MSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
pGR21                ..............MSAIKPVMKIELVMEGEVNGHKFTITGEGQGKPYEGTQTLNLTVT
consensus                          msaikpvmkielvmegevnghkftitGEGqGKPYEGTqtlnLtVt

hybridPolyth2-Disco1 KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR14                KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR15                KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR17                EGGPLRFAYDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR18                KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR19                KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR20                KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
pGR21                KGVPLPFAFDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS
consensus            kGvPLpFAfDILTTAFQYGNRAFTKYPSDIPDYFKQTFPQGYTWERTMNFEDGGVCTVNS

hybridPolyth2-Disco1 DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR14                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR15                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR17                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR18                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR19                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR20                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
pGR21                DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG
consensus            DVSLKSGCFEYKIRFDGENFPPNGPVMQKKTVKWEPSTENMYEHDGMLKGDVSRTLLLEG

hybridPolyth2-Disco1 GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSKLPSR
pGR14                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSKLPSR
pGR15                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSKLPSR
pGR17                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSKLPSR
pGR18                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSKLPSR
pGR19                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSKLPSR
pGR20                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSNLPSR
pGR21                GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSNLPSR
consensus            GGHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILSHDKDYKMVKVYEIAEAHYSkLPSR

hybridPolyth2-Disco1 LTK
pGR14                LTK
pGR15                LTK
pGR17                LTK
pGR18                LTK
pGR19                LTK
pGR20                LTK
pGR21                LTK
consensus            LTK
```

Fig 7

```
1                                   70
Translation of 3' fragment  Discosoma 1 (1)   ----------------------------------------HHMEGSVDGQNFVITGE
Translation of fragment Discosoma 2     (1)   ----------------------------------------HHMEGSVDGHYFEIKGN
Translation of PolythoalCDS             (1)   ----------------------------MSAIKPVMKVELVMEGNVNGHKFTITGE
Translation of PolythoaCDS              (1)   ----------------------------MSAIKPVMKIELVMEGEVNGHKFTITGE
Translation of WTGFP                    (1)   ----------------------------MSKGEELFTGVVPVLVELDGDVNGQKFSVSGE

                                            71                                  140
Translation of 3' fragment  Discosoma 1 (18)  GEGKPYEGTH--VIDLQVVEGGPLRFAYDILTTAFQYGNRAFTKYPSDIP--DYFKQTFPQGYTWERTMH
Translation of fragment Discosoma 2     (18)  GYGKSYDGTN--TVKLQVTKGGPLPFAWPILSPQFMYGSKAYIKHPADIA--DYKKLSFPEGFTWERVMH
Translation of PolythoalCDS             (29)  GQGKPYEGTH--TLNLTVTKGGPLPFAYDILSKAFLYGSRTFIKYPEDIP--DYFKQTFPEGYSWERTFK
Translation of PolythoaCDS              (29)  GQGKPYEGTQ--TLNLTVTKGVPLPFAFDILSTAFSYGNRVFTKYPDDIP--DYFKQTFPEGYSWERTFK
Translation of WTGFP                    (33)  GEGDATYGKL--TLNFICTTG-KLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIF
                                                                  Chromofoor

                                            141                                 210
Translation of 3' fragment  Discosoma 1 (84)  FEDGG--VCTVNSDVSLKSG---CFEYKIRFDGENFPPNG-PVMQKKTVKWEPSTEMMYE---HDG--ML
Translation of fragment Discosoma 2     (84)  FEDGG--VCCITNDISLEGN---CFIYHINFIGLNFPSDG-PV----------------------------
Translation of PolythoalCDS             (95)  Y-DEG--LCTTKSDICLKKGEPDCFQYKIYFEGKNLGPSG-PIMQKKTLKWEPSTERMYMDVDKDGAKVL
Translation of PolythoaCDS              (95)  Y-EEG--VCTTKSDISLKKGQPDCFQYKINFKGEKLDPNG-PIMQKKTLKWEPSTERMYMDVDKDGAKVL
Translation of WTGFP                    (100) YKDDG--NYKTRAEVKFEGD---TLVNRIELKGIDFKEDGNILGHKMEYNYNSHNVYIMGDKPKNG--IK

                                            211                                 280
Translation of 3' fragment  Discosoma 1 (143) KGDVSRTLLLEGG-GHYQCDFKTIYKAKDSQGIKMPEYHFVDHRIEILS-HDKDYKMVKVYEIAEAHYSK
Translation of fragment Discosoma 2     (121) ----------------------------------------------------------------------
Translation of PolythoalCDS             (161) KGDDNAALLLEGG-GHYRCDFNSIYKAKKT--GSLPAYHWIDHRIEILS-HDKDYNKVTMHEFAAARNSP
Translation of PolythoaCDS              (161) KGDVNAALLLEGG-GHYRCDFNSTYKAKKT--VSFPAYHFVDHRIEILS-HNTDYSKVTLYEVAVARNSP
Translation of WTGFP                    (163) V-NFKIRHNIKDGSVQLADHYQQNTPIGDG-PVLLPDNHYLSTQSALSKDPNEKRDHMILLEFVTAARIT

                                            281
Translation of 3' fragment  Discosoma 1 (211) LPSRLTK--
Translation of fragment Discosoma 2     (121) ---------
Translation of PolythoalCDS             (227) FPIMAPQ--
Translation of PolythoaCDS              (227) LQIMAPQ--
Translation of WTGFP                    (231) HGMDELYK-
```

Fig 8a

New Figure 8b:

```
pGR22          1 MSAIKPVMKVELVMEGNVNGHKFTITGEGQGKPYEGTHTLNLTVTKGGPLPFAYDILSAA
pGR24          1 MSAIKPVMKVELVMEGNVNGHKFTITGEGQGKPYEGTHTLNLTVTKGGPLPFAYDILSAA
pGR25          1 MSAIKPVMKVELVMEGNVNGHKFTITGEGQGKPYEGTHTLNLTVTKGGPLPFAYDILSAA
pGR26          1 MSAIKPVMKVELVMEGNVNGHKFTITGEGQGKPYEGTHTLNLTVTKGGPLPFAYDILSAA
Polythoa1      1 MSAIKPVMKVELVMEGNVNGHKFTITGEGQGKPYEGTHTLNLTVTKGGPLPFAYDILSAA

pGR22         61 FQYGNRVFTKYPEDIPDYFKQTFPEGYSWERTFKYDEGLCTTKSDICLKKGEPDCFQYKI
pGR24         61 FQYGNRVFTKYPEDIPDYFKQTFPEGYSWERTFKYDEGLCTTKSDICLKKGEPDCFQYKI
pGR25         61 FQYGNRVFTKYPEDIPDYFKQTFPEGYSWERTFKYDEGLCTTKSDICLKKGEPDCFQYKI
pGR26         61 FQYGNRVFTKYPEDIPDYFKQTFPEGYSWERTFKYDEGLCTTKSDICLKKGEPDCFQYKI
Polythoa1     61 FQYGNRVFTKYPEDIPDYFKQTFPEGYSWERTFKYDEGLCTTKSDICLKKGEPDCFQYKI

pGR22        121 YFEGKNLGPSGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDDNAALLLEGGGHYRCDF
pGR24        121 YFEGKNLGPSGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDDNAALLLEGGGHYRCDF
pGR25        121 YFEGKNLGPSGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDDNAALLLEGGGHYRCDF
pGR26        121 YFEGKNLGPSGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDDNAALLLEGGGHYRCDF
Polythoa1    121 YFEGKNLGPSGPIMQKKTLKWEPSTERMYMDVDKDGAKVLKGDDNAALLLEGGGHYRCDF

pGR22        181 NSIYKAKKTGSLPAYHWIDHRIEILSHDKDYNKVTMHEFAAARNSPFPIMAPC
pGR24        181 NSIYKAKKTGSLPAYHWIDHRIEILSHDKDYNKVTMHEFAAARNSPFPIMAPC
pGR25        181 NSIYKAKKTGSLPAYHWIDHRIEILSHDKDYNKVTMHEFAAARNSPFPIMAPC
pGR26        181 NSIYKAKKTGSLPAYHWIDHRIEILSHDKDYNKVTMHEFAAARNSPFPIMAPC
Polythoa1    181 NSIYKAKKTGSLPAYHWIDHRIEILSHDKDYNKVTMHEFAAARNSPFPIMAPC
```

Nucleotide sequence of pGR3

```
   1 aattctctac cgtcatgagt gcaattaaac cagtcatgaa gattgaattg gtcatggaag
  61 gagaggtgaa cgggcacaag ttcacgatca cgggagaggg acaaggcaag ccttacgagg
 121 gaacacagac tctagacctt acagtcacta aaggcgtgcc ccttcctttc gctttcgata
 181 tcttgtcaac agcattccag tatggcaaca gggtatttac caaataccca gatgatatac
 241 cggactattt caagcagacc tttccggaag gatattcgtg ggaaagaact ttcaaatatg
 301 aagagggcgt ttgcaccaca aagagtgaca taagcctcaa gaaaggccaa ccagactgct
 361 ttcaatataa aattaacttt aaaggggaga agcttgaccc caacggccca attatgcaga
 421 agaagaccct gaaatgggag ccatccactg agaggatgta catggacgtg gataaagacg
 481 gtgcaaaggt gctgaagggc gatgttaatg cggccctgtt gcttgaagga ggtggccatt
 541 atcgttgtga ctttaacagt acttacaagg cgaagaaaac tgtgtccttc ccagcatatc
 601 actttgtgga ccaccgcatt gagattttga gccacaatac ggattacagc aaggttacgc
 661 tgtatgaagt tgccgtggct cgcaatctcg agcaccacca ccaccaccac tgagatccgg
 721 ctgctaacaa agcccgaaag gaagctgagt tggctgctgc caccgctgag caataactag
 781 cataacccct tggggcctct aaacgggtct tgaggggttt tttgctgaaa ggaggaacta
 841 tatccggatt ggcgaatggg acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt
 901 ggttacgcgc agcgtgaccg ctacacttgc cagcgcccta gcgcccgctc ctttcgcttt
 961 cttcccttcc tttctcgcca cgttcgccgg ctttccccgt caagctctaa atcgggggct
1021 ccctttaggg ttccgattta gtgctttacg gcacctcgac cccaaaaaac ttgattaggg
1081 tgatggttca cgtagtgggc catcgccctg atagacggtt tttcgccctt tgacgttgga
1141 gtccacgttc tttaatagtg gactcttgtt ccaaactgga acaacactca accctatctc
1201 ggtctattct tttgatttat aagggatttt gccgatttcg gcctattggt taaaaaatga
1261 gctgatttaa caaaaattta acgcgaattt taacaaaata ttaacgttta caatttcagg
1321 tggcactttt cggggaaatg tgcgcggaac ccctatttgt ttatttttct aaatacattc
1381 aaatatgtat ccgctcatga gacaataacc ctgataaatg cttcaataat attgaaaaag
1441 gaagagtatg agtattcaac atttccgtgt cgcccttatt cccttttttg cggcattttg
1501 ccttcctgtt tttgctcacc cagaaacgct ggtgaaagta aaagatgctg aagatcagtt
1561 gggtgcacga gtgggttaca tcgaactgga tctcaacagc ggtaagatcc ttgagagttt
1621 tcgccccgaa gaacgttttc caatgatgag cacttttaaa gttctgctat gtggcgcggt
1681 attatcccgt attgacgccg ggcaagagca actcggtcgc cgcatacact attctcagaa
1741 tgacttggtt gagtactcac cagtcacaga aaagcatctt acggatggca tgacagtaag
1801 agaattatgc agtgctgcca taaccatgag tgataacact gcggccaact tacttctgac
1861 aacgatcgga ggaccgaagg agctaaccgc ttttttgcac aacatggggg atcatgtaac
1921 tcgccttgat cgttgggaac cggagctgaa tgaagccata ccaaacgacg agcgtgacac
1981 cacgatgcct gcagcaatgg caacaacgtt gcgcaaacta ttaactggcg aactacttac
2041 tctagcttcc cggcaacaat taatagactg gatggaggcg gataaagttg caggaccact
2101 tctgcgctcg gcccttccgg ctggctggtt tattgctgat aaatctggag ccggtgagcg
2161 tgggtctcgc ggtatcattg cagcactggg gccagatggt aagccctccc gtatcgtagt
2221 tatctacacg acggggagtc aggcaactat ggatgaacga aatagacaga tcgctgagat
2281 aggtgcctca ctgattaagc attggtaact gtcagaccaa gtttactcat atatacttta
2341 gattgattta aaacttcatt tttaatttaa aaggatctag gtgaagatcc tttttgataa
2401 tctcatgacc aaaatccctt aacgtgagtt ttcgttccac tgagcgtcag accccgtaga
2461 aaagatcaaa ggatcttctt gagatccttt ttttctgcgc gtaatctgct gcttgcaaac
2521 aaaaaaacca ccgctaccag cggtggtttg tttgccggat caagagctac caactctttt
2581 tccgaaggta actggcttca gcagagcgca gataccaaat actgtccttc tagtgtagcc
2641 gtagttaggc caccacttca agaactctgt agcaccgcct acatacctcg ctctgctaat
2701 cctgttacca gtggctgctg ccagtggcga taagtcgtgt cttaccgggt tggactcaag
2761 acgatagtta ccggataagg cgcagcggtc gggctgaacg gggggttcgt gcacacagcc
2821 cagcttggag cgaacgacct acaccgaact gagataccta cagcgtgagc tatgagaaag
2881 cgccacgctt cccgaaggga gaaaggcgga caggtatccg gtaagcggca gggtcggaac
2941 aggagagcgc acgagggagc ttccaggggg aaacgcctgg tatctttata gtcctgtcgg
3001 gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct
3061 atggaaaaac gccagcaacg cggccttttt acggttcctg gccttttgct ggccttttgc
3121 tcacatgttc tttcctgcgt tatcccctga ttctgtggat aaccgtatta ccgcctttga
3181 gtgagctgat accgctcgcc gcagccgaac gaccgagcgc agcgagtcag tgagcgagga
3241 agcggaagag cgcctgatgc ggtattttct ccttacgcat ctgtgcggta tttcacaccg
3301 catatatggt gcactctcag tacaatctgc tctgatgccg catagttaag ccagtataca
3361 ctccgctatc gctacgtgac tgggtcatgg ctgcgccccg acacccgcca acacccgctg
3421 acgcgccctg acgggcttgt ctgctcccgg catccgctta cagacaagct gtgaccgtct
```

Fig 9

Fig 9 (cont'd)

```
3481 ccgggagctg catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg aggcagctgc
3541 ggtaaagctc atcagcgtgg tcgtgaagcg attcacagat gtctgcctgt tcatccgcgt
3601 ccagctcgtt gagtttctcc agaagcgtta atgtctggct tctgataaag cgggccatgt
3661 taagggcggt tttttcctgt ttggtcactg atgcctccgt gtaaggggga tttctgttca
3721 tgggggtaat gataccgatg aaacgagaga ggatgctcac gatacgggtt actgatgatg
3781 aacatgcccg gttactggaa cgttgtgagg gtaaacaact ggcggtatgg atgcggcggg
3841 accagagaaa aatcactcag ggtcaatgcc agcgcttcgt taatacagat gtaggtgttc
3901 cacagggtag ccagcagcat cctgcgatgc agatccggaa cataatggtg cagggcgctg
3961 acttccgcgt ttccagactt tacgaaacac ggaaaccgaa gaccattcat gttgttgctc
4021 aggtcgcaga cgttttgcag cagcagtcgc ttcacgttcg ctcgcgtatc ggtgattcat
4081 tctgctaacc agtaaggcaa ccccgccagc ctagccgggt cctcaacgac aggagcacga
4141 tcatgcgcac ccgtggggcc gccatgccgg cgataatggc ctgcttctcg ccgaaacgtt
4201 tggtggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg aataccgcaa
4261 gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa atgacccaga
4321 gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata agtgcggcga
4381 cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct ctcaagggca
4441 tcggtcgaga tcccggtgcc taatgagtga gctaacttac attaattgcg ttgcgctcac
4501 tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg
4561 cggggagagg cggtttgcgt attgggcgcc agggtggttt ttcttttcac cagtgagacg
4621 ggcaacagct gattgccctt caccgcctgg ccctgagaga gttgcagcaa gcggtccacg
4681 ctggtttgcc ccagcaggcg aaaatcctgt ttgatggtgg ttaacggcgg gatataacat
4741 gagctgtctt cggtatcgtc gtatcccact accgagatgt ccgcaccaac gcgcagcccg
4801 gactcggtaa tggcgcgcat tgcgcccagc gccatctgat cgttggcaac cagcatcgca
4861 gtgggaacga tgccctcatt cagcatttgc atggtttgtt gaaaaccgga catggcactc
4921 cagtcgcctt cccgttccgc tatcggctga atttgattgc gagtgagata tttatgccag
4981 ccagccagac gcagacgcgc cgagacagaa cttaatgggc ccgctaacag cgcgatttgc
5041 tggtgaccca atgcgaccag atgctccacg cccagtcgcg taccgtcttc atgggagaaa
5101 ataatactgt tgatgggtgt ctggtcagag acatcaagaa ataacgccgg aacattagtg
5161 caggcagctt ccacagcaat ggcatcctgg tcatccagcg gatagttaat gatcagccca
5221 ctgacgcgtt gcgcgagaag attgtgcacc gccgctttac aggcttcgac gccgcttcgt
5281 tctaccatcg acaccaccac gctggcaccc agttgatcgg cgcgagattt aatcgccgcg
5341 acaatttgcg acggcgcgtg cagggccaga ctggaggtgg caacgccaat cagcaacgac
5401 tgtttgcccg ccagttgttg tgccacgcgg ttgggaatgt aattcagctc cgccatcgcc
5461 gcttccactt tttcccgcgt tttcgcagaa acgtggctgg cctggttcac cacgcgggaa
5521 acggtctgat aagagacacc ggcatactct gcgacatcgt ataacgttac tggtttcaca
5581 ttcaccaccc tgaattgact ctcttccggg cgctatcatg ccataccgcg aaaggttttg
5641 cgccattcga tggtgtccgg gatctcgacg ctctcccttа tgcgactcct gcattaggaa
5701 gcagcccagt agtaggttga ggccgttgag caccgccgcc gcaaggaatg gtgcatgcaa
5761 ggagatggcg cccaacagtc ccccggccac ggggcctgcc accataccca cgccgaaaca
5821 agcgctcatg agcccgaagt ggcgagcccg atcttcccca tcggtgatgt cggcgatata
5881 ggcgccagca accgcacctg tggcgccggt gatgccggcc acgatgcgtc cggcgtagag
5941 gatcgagatc gatctcgatc ccgcgaaatt aatacgactc actataggg aattgtgagc
6001 ggataacaat tcccctctag aaataatttt gtttaacttt aagaaggaga tatacatatg
6061 agcgataaaa ttattcacct gactgacgac agttttgaca cggatgtact caaagcggac
6121 ggggcgatcc tcgtcgattt ctgggcagag tggtgcggtc cgtgcaaaat gatcgccccg
6181 attctggatg aaatcgctga cgaatatcag ggcaaactga ccgttgcaaa actgaacatc
6241 gatcaaaacc ctggcactgc gccgaaatat ggcatccgtg gtatcccgac tctgctgctg
6301 ttcaaaaacg gtgaagtggc ggcaaccaaa gtgggtgcac tgtctaaagg tcagttgaaa
6361 gagttcctcg acgctaacct ggccggttct ggttctggcc atatgcacca tcatcatcat
6421 cattcttctg gtctggtgcc acgcggttct ggtatgaaag aaaccgctgc tgctaaattc
6481 gaacgccagc acatggacag cccagatctg ggtaccgacg acgacgacaa ggccatggct
6541 gatatcggat ccg
```

nucleotide sequence of pGR4

```
1 aattctctac cgtcatgagt gcaattaaac cagtcatgaa gattgaattg gtcatggaag
  61 gagaggtgaa cgggcacaag ttcacgatca cgggagaggg acaaggcaag ccttacgagg
 121 gaacacagac tctagacctt acagtcacta aaggcgtgcc ccttcctttc gctttcgata
 181 tcttgtcaac agcattccag tatggcaaca gggtatttac caaataccca gatgatatac
 241 cggactattt caagcagacc tttccggaag gatattcgtg ggaaagaact ttcaaatatg
 301 aagagggcgt ttgcaccaca aagagtgaca taagcctcaa gaaaggccaa ccagactgct
 361 ttcaatataa aattaacttt aaaggggaga agcttgaccc caacggccca attatgcaga
```

Fig 10

```
 421 agaagaccct gaaatgggag ccatccactg agaggatgta catggacgtg gataaagacg
 481 gtgcaaaggt gctgaagggc gatgttaatg cggccctgtt gcttgaagga ggtggccatt
 541 atcgttgtga ctttaacagt acttacaagg cgaagaaaac tgtgtccttc ccagcatatc
 601 actttgtgga ccaccgcatt gagattttga gccacaatac ggattacagc aaggttacgc
 661 tgtatgaagt tgccgtggct cgcaatctcg agcatgcatc tagagggccc tattctatag
 721 tgtcacctaa atgctagagc tcgctgatca gcctcgactg tgccttctag ttgccagcca
 781 tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac tcccactgtc
 841 ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca ttctattctg
 901 gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag caggcatgct
 961 ggggatgcgg tgggctctat ggcttctgag gcggaaagaa ccagctgggg ctctaggggg
1021 tatccccacg cgccctgtag cggcgcatta agcgcggcgg gtgtggtggt tacgcgcagc
1081 gtgaccgcta cacttgccag cgccctagcg cccgctcctt tcgctttctt cccttcctttt
1141 ctcgccacgt tcgccggctt tccccgtcaa gctctaaatc ggggcatccc tttagggttc
1201 cgatttagtg ctttacggca cctcgacccc aaaaaacttg attagggtga tggttcacgt
1261 agtgggccat cgccctgata gacggttttt cgccctttga cgttggagtc cacgttcttt
1321 aatagtggac tcttgttcca aactggaaca acactcaacc ctatctcggt ctattctttt
1381 gatttataag ggattttggg gatttcggcc tattggttaa aaaatgagct gatttaacaa
1441 aaatttaacg cgaattaatt ctgtggaatg tgtgtcagtt agggtgtgga aagtccccag
1501 gctccccagg caggcagaag tatgcaaagc atgcatctca attagtcagc aaccaggtgt
1561 ggaaagtccc caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca
1621 gcaaccatag tcccgccccc aactccgccc atcccgcccc taactccgcc cagttccgcc
1681 cattctccgc cccatggctg actaatttttt tttatttatg cagaggccga ggccgcctct
1741 gcctctgagc tattccagaa gtagtgagga ggcttttttg gaggcctagg cttttgcaaa
1801 aagctcccgg gagcttgtat atccattttc ggatctgatc aagagacagg atgaggatcg
1861 tttcgcatga ttgaacaaga tggattgcac gcaggttctc cggccgcttg ggtggagagg
1921 ctattcggct atgactgggc acaacagaca atcggctgct ctgatgccgc cgtgttccgg
1981 ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg acctgtccgg tgccctgaat
2041 gaactgcagg acgaggcagc gcggctatcg tggctggcca cgacgggcgt tccttgcgca
2101 gctgtgctcg acgttgtcac tgaagcggga agggactggc tgctattggg cgaagtgccg
2161 gggcaggatc tcctgtcatc tcaccttgct cctgccgaga aagtatccat catggctgat
2221 gcaatgcggc ggctgcatac gcttgatccg gctacctgcc cattcgacca ccaagcgaaa
2281 catcgcatcg agcgagcacg tactcggatg gaagccggtc ttgtcgatca ggatgatctg
2341 gacgaagagc atcaggggct cgcgccagcc gaactgttcg ccaggctcaa ggcgcgcatg
2401 cccgacggcg aggatctcgt cgtgacccat ggcgatgcct gcttgccgaa tatcatggtg
2461 gaaaatggcc gcttttctgg attcatcgac tgtggccggc tgggtgtggc ggaccgctat
2521 caggacatag cgttggctac ccgtgatatt gctgaagagc ttggcggcga atgggctgac
2581 cgcttcctcg tgctttacgg tatcgccgct cccgattcgc agcgcatcgc cttctatcgc
2641 cttcttgacg agttcttctg agcgggactc tggggttcga aatgaccgac caagcgacgc
2701 ccaacctgcc atcacgagat ttcgattcca ccgccgcctt ctatgaaagg ttgggcttcg
2761 gaatcgtttt ccgggacgcc ggctggatga tcctccagcg cggggatctc atgctggagt
2821 tcttcgccca ccccaacttg tttattgcag cttataatgg ttacaaataa agcaatagca
2881 tcacaaattt cacaaataaa gcatttttttt cactgcattc tagttgtggt ttgtccaaac
2941 tcatcaatgt atcttatcat gtctgtatac cgtcgacctc tagctagagc ttggcgtaat
3001 catggtcata gctgtttcct gtgtgaaatt gttatccgct cacaattcca cacaacatac
3061 gagccggaag cataaagtgt aaagcctggg gtgcctaatg agtgagctaa ctcacattaa
3121 ttgcgttgcg ctcactgccc gctttccagt cgggaaacct gtcgtgccag ctgcattaat
3181 gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc gcttcctcgc
3241 tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg
3301 cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag
3361 gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc
3421 gcccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag
3481 gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga
3541 ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc
3601 aatgctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg
3661 tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt
3721 ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca
3781 gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca
3841 ctagaaggac agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag
3901 ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca
3961 agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg
4021 ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa
4081 aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta
```

Fig 10 (cont'd)

```
4141 tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag
4201 cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga
4261 tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac ccacgctcac
4321 cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc agaagtggtc
4381 ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct agagtaagta
4441 gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac
4501 gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat
4561 gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc gttgtcagaa
4621 gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat tctcttactg
4681 tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag tcattctgag
4741 aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat aataccgcgc
4801 cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct
4861 caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat
4921 cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg
4981 ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc ttcctttttc
5041 aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta
5101 tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg
5161 tcgacggatc gggagatctc ccgatcccct atggtcgact ctcagtacaa tctgctctga
5221 tgccgcatag ttaagccagt atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg
5281 cgcgagcaaa atttaagcta caacaaggca aggcttgacc gacaattgca tgaagaatct
5341 gcttagggtt aggcgttttg cgctgcttcg cgatgtacgg gccagatata cgcgttgaca
5401 ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc atagcccata
5461 tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac cgcccaacga
5521 cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa tagggacttt
5581 ccattgacgt caatgggtgg actatttacg gtaaactgcc cacttggcag tacatcaagt
5641 gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgcctggca
5701 ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct acgtattagt
5761 catcgctatt accatggtga tgcggttttg gcagtacatc aatgggcgtg gatagcggtt
5821 tgactcacgg ggatttccaa gtctccaccc cattgacgtc aatgggagtt tgttttggca
5881 ccaaaatcaa cgggactttc caaaatgtcg taacaactcc gccccattga cgcaaatggg
5941 cggtaggcgt gtacggtggg aggtctatat aagcagagct ctctggctaa ctagagaacc
6001 cactgcttac tggcttatcg aaattaatac gactcactat agggagaccc aagcttggta
6061 ccgagctcgg atccactagt aacggccgcc agtgtgctgg
```

Fig 10 (cont'd)

nucleotide sequence of pGR5

```
1 aattcgccct tctggaattc tttaccgtca tgagtgcaat taaaccagtc atgaagattg
  61 aattggtcat ggaaggagag gtgaacgggc acaagttcac gatcacggga gagggacaag
 121 gcaagcctta cgagggaaca cagactctag accttacagt cactaaaggc gtgccccttc
 181 ctttcgcttt cgatatcttg tcaacagcat tccagtatgg caacagggta tttaccaaat
 241 acccagatga tataccggac tatttcaagc agacctttcc ggaaggatat tcgtgggaaa
 301 gaactttcaa atatgaagag ggcgtttgca ccacaaagag tgacataagc ctcaagaaag
 361 gccaaccaga ctgctttcaa tataaaatta actttaaagg ggagaagctt gaccccaacg
 421 gcccaattat gcagaagaag accctgaaat gggagccatc cactgagagg atgtacatgg
 481 acgtggataa agacggtgca aaggtgctga agggcgatgt taatgcggcc ctgttgcttg
 541 aaggaggtgg ccattatcgt tgtgacttta acagtactta caaggcgaag aaaactgtgt
 601 ccttcccagc atatcacttt gtggaccacc gcattgagat tttgagccac aatacggatt
 661 acagcaaggt tacactgtat gaagttgccg tggctcgcaa ttctcctctt cagattatag
 721 cctcgagcat gcatctagag ggccctattc tatagtgtca cctaaatgct agagctcgct
 781 gatcagcctc gactgtgcct tctagttgcc agccatctgt tgtttgcccc tcccccgtgc
 841 cttccttgac cctggaaggt gccactccca ctgtcctttc ctaataaaat gaggaaattg
 901 catcgcattg tctgagtagg tgtcattcta ttctgggggg tggggtgggg caggacagca
 961 agggggagga ttgggaagac aatagcaggc atgctgggga tgcggtgggc tctatggctt
1021 ctgaggcgga aagaaccagc tggggctcta gggggtatcc ccacgcgccc tgtagcggcg
1081 cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcgccc
1141 tagcgcccgc tcctttcgct ttcttcccct cctttctcgc cacgttcgcc ggctttcccc
1201 gtcaagctct aaatcggggc atcccctttag ggttccgatt tagtgcttta cggcacctcg
1261 accccaaaaa acttgattag ggtgatggtt cacgtagtgg gccatcgccc tgatagacgg
1321 tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg ttccaaactg
1381 gaacaacact caaccctatc tcggtctatt cttttgattt ataagggatt ttggggattt
1441 cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat taattctgtg
1501 gaatgtgtgt cagttagggt gtggaaagtc cccaggctcc ccaggcaggc agaagtatgc
```

Fig 11

```
1561 aaagcatgca tctcaattag tcagcaacca ggtgtggaaa gtccccaggc tccccagcag
1621 gcagaagtat gcaaagcatg catctcaatt agtcagcaac catagtcccg cccctaactc
1681 cgcccatccc gcccctaact ccgcccagtt ccgcccattc tccgccccat ggctgactaa
1741 ttttttttat ttatgcagag gccgaggccg cctctgcctc tgagctattc cagaagtagt
1801 gaggaggctt ttttggaggc ctaggctttt gcaaaaagct cccgggagct tgtatatcca
1861 ttttcggatc tgatcaagag acaggatgag gatcgtttcg catgattgaa caagatggat
1921 tgcacgcagg ttctccggcc gcttgggtgg agaggctatt cggctatgac tgggcacaac
1981 agacaatcgg ctgctctgat gccgccgtgt tccggctgtc agcgcagggg cgcccggttc
2041 tttttgtcaa gaccgacctg tccggtgccc tgaatgaact gcaggacgag gcagcgcggc
2101 tatcgtggct ggccacgacg ggcgttcctt gcgcagctgt gctcgacgtt gtcactgaag
2161 cgggaaggga ctggctgcta ttgggcgaag tgccggggca ggatctcctg tcatctcacc
2221 ttgctcctgc cgagaaagta tccatcatgg ctgatgcaat gcggcggctg catacgcttg
2281 atccggctac ctgcccattc gaccaccaag cgaaacatcg catcgagcga gcacgtactc
2341 ggatggaagc cggtcttgtc gatcaggatg atctggacga agagcatcag gggctcgcgc
2401 cagccgaact gttcgccagg ctcaaggcgc gcatgcccga cggcgaggat ctcgtcgtga
2461 cccatggcga tgcctgcttg ccgaatatca tggtggaaaa tggccgcttt tctggattca
2521 tcgactgtgg ccggctgggt gtggcggacc gctatcagga catagcgttg gctacccgtg
2581 atattgctga agagcttggc ggcgaatggg ctgaccgctt cctcgtgctt tacggtatcg
2641 ccgctcccga ttcgcagcgc atcgccttct atcgccttct tgacgagttc ttctgagcgg
2701 gactctgggg ttcgaaatga ccgaccaagc gacgcccaac ctgccatcac gagatttcga
2761 ttccaccgcc gccttctatg aaaggttggg cttcggaatc gttttccggg acgccggctg
2821 gatgatcctc cagcgcgggg atctcatgct ggagttcttc gcccacccca acttgtttat
2881 tgcagcttat aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt
2941 tttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt atcatgtctg
3001 tataccgtcg acctctagct agagcttggc gtaatcatgg tcatagctgt ttcctgtgtg
3061 aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa agtgtaaagc
3121 ctggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac tgcccgcttt
3181 ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg
3241 cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc gctcggtcgt
3301 tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat ccacagaatc
3361 aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa
3421 aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa
3481 tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc
3541 ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc
3601 cgcctttctc ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta ggtatctcag
3661 ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaacccccgg ttcagcccga
3721 ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc
3781 gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac
3841 agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg
3901 cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca
3961 aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa
4021 aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa
4081 ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt
4141 aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag
4201 ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat
4261 agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc
4321 cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa
4381 ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca
4441 gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa
4501 cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta tggcttcatt
4561 cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc
4621 ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact
4681 catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc
4741 tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg
4801 ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgct
4861 catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc
4921 cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatcttta ctttcaccag
4981 cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac
5041 acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg
5101 ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac aaatagggg
5161 tccgcgcaca tttccccgaa aagtgccacc tgacgtcgac ggatcgggag atctcccgat
5221 cccctatggt cgactctcag tacaatctgc tctgatgccg catagttaag ccagtatctg
```

Fig 11 (cont'd)

```
5281 ctccctgctt gtgtgttgga ggtcgctgag tagtgcgcga gcaaaattta agctacaaca
5341 aggcaaggct tgaccgacaa ttgcatgaag aatctgctta gggttaggcg ttttgcgctg
5401 cttcgcgatg tacgggccag atatacgcgt tgacattgat tattgactag ttattaatag
5461 taatcaatta cggggtcatt agttcatagc ccatatatgg agttccgcgt tacataactt
5521 acggtaaatg gcccgcctgg ctgaccgccc aacgacccca gcccattgac gtcaataatg
5581 acgtatgttc ccatagtaac gccaatagggg actttccatt gacgtcaatg ggtggactat
5641 ttacggtaaa ctgcccactt ggcagtacat caagtgtatc atatgccaag tacgcccccc
5701 attgacgtca atgacggtaa atggcccgcc tggcattatg cccagtacat gaccttatgg
5761 gactttccta cttggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg
5821 ttttggcagt acatcaatgg gcgtggatag cggtttgact cacggggatt tccaagtctc
5881 caccccattg acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa
5941 tgtcgtaaca actccgcccc attgacgcaa atgggcggta ggcgtgtacg gtgggaggtc
6001 tatataagca gagctctctg gctaactaga gaacccactg cttactggct tatcgaaatt
6061 aatacgactc actataggga gacccaagct tggtaccgag ctcggatcca ctagtaacgg
6121 ccgccagtgt gctgg
```

Fig 11 (cont'd)

Nucleotide sequence of pGR6

```
1 tcgagatgca tggccggccg agctccgcat cggccgctgt catcagatcg ccatctcgcg
   61 cccgtgcctc tgacttctaa gtccaattac tcttcaacat ccctacatgc tctttctccc
  121 tgtgctccca ccccctattt ttgttattat caaaaaaact tcttcttaat ttctttgttt
  181 tttagcttct tttaagtcac ctctaacaat gaaattgtgt agattcaaaa atagaattaa
  241 ttcgtaataa aaagtcgaaa aaaattgtgc tccctccccc cattaataat aattctatcc
  301 caaaatctac acaatgttct gtgtacactt cttatgtttt ttttacttct gataaatttt
  361 ttttgaaaca tcatagaaaa aaccgcacac aaaatacctt atcatatgtt acgtttcagt
  421 ttatgaccgc aatttttatt tcttcgcacg tctgggcctc tcatgacgtc aaatcatgct
  481 catcgtgaaa aagttttgga gtatttttgg aatttttcaa tcaagtgaaa gtttatgaaa
  541 ttaattttcc tgcttttgct ttttgggggt ttcccctatt gtttgtcaag agtttcgagg
  601 acggcgtttt tcttgctaaa atcacaagta ttgatgagca cgatgcaaga aagatcggaa
  661 gaaggtttgg gtttgaggct cagtggaagg tgagtagaag ttgataattt gaaagtggag
  721 tagtgtctat ggggtttttg ccttaaatga cagaatacat tcccaatata ccaaacataa
  781 ctgtttccta ctagtcggcc gtacgggccc tttcgtctcg cgcgtttcgg tgatgacggt
  841 gaaaacctct gacacatgca gctcccggag acggtcacag cttgtctgta agcggatgcc
  901 gggagcagac aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg gggctggctt
  961 aactatgcgg catcagagca gattgtactg agagtgcacc atatgcggtg tgaaataccg
 1021 cacagatgcg taaggagaaa ataccgcatc aggcggcctt aagggcctcg tgatacgcct
 1081 atttttatag gttaatgtca tgataataat ggtttcttag acgtcaggtg gcacttttcg
 1141 gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa atatgtatcc
 1201 gctcatgaga caataaccct gataaatgct tcaataatat tgaaaaagga agagtatgag
 1261 tattcaacat ttccgtgtcg cccttattcc ctttttgcg gcattttgcc ttcctgtttt
 1321 tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt
 1381 gggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc gccccgaaga
 1441 acgtttttcca atgatgagca cttttaaagt tctgctatgt ggcgcggtat tatcccgtat
 1501 tgacgccggg caagagcaac tcggtcgccg catacactat tctcagaatg acttggttga
 1561 gtactcacca gtcacagaaa agcatcttac ggatggcatg acagtaagag aattatgcag
 1621 tgctgccata accatgagtg ataacactgc ggccaactta cttctgacaa cgatcggagg
 1681 accgaaggag ctaaccgctt ttttgcacaa catgggggat catgtaactc gccttgatcg
 1741 ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca cgatgcctgt
 1801 agcaatggca acaacgttgc gcaaactatt aactggcgaa ctacttactc tagcttcccg
 1861 gcaacaatta atagactgga tggaggcgga taaagttgca ggaccacttc tgcgctcggc
 1921 ccttccggct ggctggttta ttgctgataa atctggagcc ggtgagcgtg ggtctcgcgg
 1981 tatcattgca gcactggggc cagatggtaa gccctcccgt atcgtagtta tctacacgac
 2041 ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag gtgcctcact
 2101 gattaagcat tggtaactgt cagaccaagt ttactcatat atactttaga ttgatttaaa
 2161 acttcatttt taatttaaaa ggatctaggt gaagatcctt tttgataatc tcatgaccaa
 2221 aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg
 2281 atcttcttga tatcctttttt ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc
 2341 gctaccagcg gtggtttgtt tgccggatca agagctacca actcttttttc cgaaggtaac
 2401 tggcttcagc agagcgcaga taccaaatac tgtccttcta gtgtagccgt agttaggcca
 2461 ccacttcaag aactctgtag caccgcctac atacctcgct ctgctaatcc tgttaccagt
 2521 ggctgctgcc agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc
 2581 ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca gcttggagcg
```

Fig 12

Fig 12 (cont'd)

```
2641 aacgacctac accgaactga gatacctaca gcgtgagcat tgagaaagcg ccacgcttcc
2701 cgaagggaga aaggcggaca ggtatccggt aagcggcagg gtcggaacag gagagcgcac
2761 gagggagctt ccagggggaa acgcctggta tctttatagt cctgtcgggt ttcgccacct
2821 ctgacttgag cgtcgatttt tgtgatgctc gtcaggggcg cggagcctat ggaaaaacgc
2881 cagcaacgcg gccttttac ggttcctggc cttttgctgg ccttttgctc acatgttctt
2941 tcctgcgtta tcccctgatt ctgtggataa ccgtattacc gcctttgagt gagctgatac
3001 cgctcgccgc agccgaacga ccgagcgcag cgagtcagtg agcgaggaag cggaagagcg
3061 cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt cattaatgca gctggcacga
3121 caggtttccc gactggaaag cgggcagtga gcgcaacgca attaatgtga gttagctcac
3181 tcattaggca ccccaggctt tacactttat gcttccggct cgtatgttgt gtggaattgt
3241 gagcggataa caatttcaca caggaaacag ctatgaccat gattacgcca agcttgcatg
3301 cctgcaggtc gactctagag gatccccagc ttgcatgcct gcaggtcgag gcatttgaat
3361 tgggggtggt ggacagtaac tgtctgtaat aataattact cctgaccagg ttgcaattcg
3421 agttttgata agcataatta taccttgtac attgtgggtt ttgtgctgtg gacgttttat
3481 tgtggacatc cccataagct acaagaaacc aaaaatgaaa ttaaaagtat tgaaaaacgt
3541 cgtaacattt tatatctgag tagtatcctt tgctttaaat gtccataaaa ataattttat
3601 aatcaataaa acaacgtttg taaatcaact gagtttacaa gtagagacat tgagggatac
3661 tttcactatg ctaaagtgaa taatcgacca aataataact cactttggta tttattcctg
3721 tcttataatg ttatgtatga attaaattca tatgcatatg gctcactctg acaaaaaaaa
3781 ataatcttcc agatcaatat tgactaccga tgcgggtggt cttttgcttt gaattctgct
3841 gaactttaca ccccgaacag caatgtgtgc ttcagctaaa aaaaagtaag tgtgttaatc
3901 agtccccccg attcttcatt ttttgccccc ctctcccgtt tcgtcggcaa aagaagagaa
3961 aataaagata agtctcaaga taggttggta atcgctaaag tggttgtgtg gataagagta
4021 gcaaaatggc aggaagagca ctttgcgcgc acacactgta ctcattgttc tggataaaat
4081 tctctcgttg tttgccgtcg gatgtctgcc tctctgccat tgagccggct tctttcactat
4141 ctttagttaa cctaaaatgc cgtttctttt ctcgtatccc actatccgtt gaggttctct
4201 gctctcttcg ctcccttacc gccagcgagc aactatccgt gggggcgcct tgctcggaag
4261 atgggggggga agaaagaaga tttttgctat ttgcacttga gaaagagact tttcctgcgt
4321 cgatggttag agaacagtgt gcagacactt ttcagctacc tagatacatg gatatccccg
4381 cctcccaatc cacccaccca gggaaaaaga agggctcgcc gaaaaatcaa agttatctcc
4441 aggctcgcgc atcccaccga gcggttgact tctctccacc acttttcatt ttaaccctcg
4501 gggtacggga ttggccaaag gacccaaagg tatgtttcga atgatactaa cataacatag
4561 aacattttca ggaggaccct tggctagcgt cgacggtacc atggggcgcg ccaccaccat
4621 gagtgcaatt aagccagtta tgaagattga attggtcatg gaaggagagg tgaacgggca
4681 caagttcacg atcacgggag agggacaagg caagccttac gagggaacac agactctaaa
4741 ccttacagtc actaaaggcg tgccccttcc tttcgctttc gatatcttgt caacagcatt
4801 ccagtatggc aacagggtat ttaccaaata cccagatgat ataccggact atttcaagca
4861 gacctttccg gaaggatatt cgtgggaaag aactttcaaa tatgaagagg gcgtttgcac
4921 cacaaagagt gacataagcc tcaagaaagg ccaaccagac tgctttcaat ataaaattaa
4981 ctttaaaggg gagaagcttg accccaacgg cccaattatg cagaagaaga ccctgaaatg
5041 ggagccatcc actgagagga tgtacatgga tgtggataaa gacggtgcaa aggtgctgaa
5101 gggcgatgtt aatgcggccc tgttgcttga aggaggcggc cattatcgtt gtgactttaa
5161 cagtacttac aaggcgaaga aaactgtgtc cttcccagca tatcactttg tggaccaccg
5221 cattgagatt ttgagccaca atacggatta cagcaaggtt acactgtatg aagttgccgt
5281 ggctcgcaat tctcctcttc agattatggc gccccagtaa aggcttaacg aaaagccaag
5341 actc
```

Nucleotide sequence of pGR7

```
1 aattctatta ctttgagtct accatcatga gtgcaattaa accagtcatg aagattgaat
  61 tggtcatgga aggagaggtg aacgggcaca agttcacgat cacgggagag ggacaaggca
 121 agccttacga gggaacacag actctaaacc ttacagtcac taaaggcgtg ccccttcctt
 181 tcgctttcga tatcttgtca acagcattcc agtatggcaa cagggtattt accaaatacc
 241 cagatgatat accggactat ttcaagcaga cctttccgga aggatattcg tgggaaagaa
 301 ctttcaaata tgaagagggc gtttgcacca caaagagtga cataagcctc aagaaaggcc
 361 aaccagactg ctttcaatat aaaattaact ttaaagggga gaagcttgac cccaacggcc
 421 caattatgcg gaagaagacc ctgaaatggg agccatccac tgagaggatg tacatggatg
 481 tggataaaga cggtgcaaag gtgctgaagg gcgatgttaa tgcggccctg ttgcttgaag
 541 gaggcggcca ttatcgttgt gactttaaca gtacttacaa ggcgaagaaa actgtgtcct
 601 tcccagcata tcactttgtg gaccaccgca ttgagatttt gagccacaat acggattaca
 661 gcaaggttac actgtatgaa gttgccgtgg ctcgcaattc tcctcttcag attatggcgc
 721 cccagtaaag gcttaacgaa aagccaagac gctcgagcac caccaccacc accactgaga
```

Fig 13

781 tccggctgct aacaaagccc gaaaggaagc tgagttggct gctgccaccg ctgagcaata
841 actagcataa ccccttgggg cctctaaacg ggtcttgagg ggtttttgc tgaaaggagg
901 aactatatcc ggattggcga atgggacgcg ccctgtagcg gcgcattaag cgcggcgggt
961 gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc cgctcctttc
1021 gctttcttcc cttcctttct cgccacgttc gccggctttc cccgtcaagc tctaaatcgg
1081 gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa aaaacttgat
1141 tagggtgatg gttcacgtag tgggccatcg ccctgataga cggtttttcg ccctttgacg
1201 ttggagtcca cgttctttaa tagtggactc ttgttccaaa ctggaacaac actcaaccct
1261 atctcggtct attcttttga tttataaggg attttgccga tttcggccta ttggttaaaa
1321 aatgagctga tttaacaaaa atttaacgcg aattttaaca aaatattaac gtttacaatt
1381 tcaggtggca cttttcgggg aaatgtgcgc ggaacccctа tttgtttatt tttctaaata
1441 cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca ataatattga
1501 aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattcccct ttttgcggca
1561 ttttgccttc ctgtttttgc tcacccagaa acgctggtga aagtaaaaga tgctgaagat
1621 cagttgggtg cacgagtggg ttacatcgaa ctggatctca acagcggtaa gatccttgag
1681 agttttcgcc ccgaagaacg ttttccaatg atgagcactt ttaaagttct gctatgtggc
1741 gcggtattat cccgtattga cgccgggcaa gagcaactcg gtcgccgcat acactattct
1801 cagaatgact tggttgagta ctcaccagtc acagaaaagc atcttacgga tggcatgaca
1861 gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcggc caacttactt
1921 ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat gggggatcat
1981 gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt
2041 gacaccacga tgcctgcagc aatggcaaca acgttgcgca aactattaac tggcgaacta
2101 cttactctag cttcccggca acaattaata gactggatgg aggcggataa agttgcagga
2161 ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc tggagccggt
2221 gagcgtgggt ctcgcggtat cattgcagca ctggggccag atggtaagcc ctcccgtatc
2281 gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag acagatcgct
2341 gagataggtg cctcactgat taagcattgg taactgtcag accaagttta ctcatatata
2401 ctttagattg atttaaaact tcatttttaa tttaaaagga tctaggtgaa gatccttttt
2461 gataatctca tgaccaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc
2521 gtagaaaaga tcaaaggatc ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg
2581 caaacaaaaa aaccaccgct accagcggtg gtttgtttgc cggatcaaga gctaccaact
2641 ctttttccga aggtaactgg cttcagcaga gcgcagatac caaatactgt ccttctagtg
2701 tagccgtagt taggccacca cttcaagaac tctgtagcac cgcctacata cctcgctctg
2761 ctaatcctgt taccagtggc tgctgccagt ggcgataagt cgtgtcttac cgggttggac
2821 tcaagacgat agttaccgga taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca
2881 cagcccagct tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga
2941 gaaagcgcca cgcttcccga agggagaaag gcggacaggt atccggtaag cggcagggtc
3001 ggaacaggag agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtcct
3061 gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc agggggggcgg
3121 agcctatgga aaaacgccag caacgcggcc tttttacggt tcctggcctt ttgctggcct
3181 tttgctcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc
3241 tttgagtgag ctgataccgc tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc
3301 gaggaagcgg aagagcgcct gatgcggtat tttctcctta cgcatctgtg cggtatttca
3361 caccgcatat atggtgcact ctcagtacaa tctgctctga tgccgcatag ttaagccagt
3421 atacactccg ctatcgctac gtgactgggt catggctgcg ccccgacacc cgccaacacc
3481 cgctgacgcg ccctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac
3541 cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac gcgcgaggca
3601 gctgcggtaa agctcatcag cgtggtcgtg aagcgattca cagatgtctg cctgttcatc
3661 cgcgtccagc tcgttgagtt tctccagaag cgttaatgtc tggcttctga taaagcgggc
3721 catgttaagg gcggtttttt cctgtttggt cactgatgcc tccgtgtaag ggggatttct
3781 gttcatgggg gtaatgatac cgatgaaacg agagaggatg ctcacgatac gggttactga
3841 tgatgaacat gcccggttac tggaacgttg tgagggtaaa caactggcgg tatggatgcg
3901 gcgggaccag agaaaaatca ctcagggtca atgccagcgc ttcgttaata cagatgtagg
3961 tgttccacag ggtagccagc agcatcctgc gatgcagatc cggaacataa tggtgcaggg
4021 cgctgacttc cgcgtttcca gactttacga aacacggaaa ccgaagacca ttcatgttgt
4081 tgctcaggtc gcagacgttt tgcagcagca gtcgcttcac gttcgctcgc gtatcggtga
4141 ttcattctgc taaccagtaa ggcaaccccg ccagcctagc cgggtcctca acgacaggag
4201 cacgatcatg cgcacccgtg gggccgccat gccggcgata atggcctgct tctcgccgaa
4261 acgtttggtg gcgggaccag tgacgaaggc ttgagcgagg gcgtgcaaga ttccgaatac
4321 cgcaagcgac aggccgatca tcgtcgcgct ccagcgaaag cggtcctcgc cgaaaatgac
4381 ccagagcgct gccggcacct gtcctacgag ttgcatgata aagaagacag tcataagtgc
4441 ggcgacgata gtcatgcccc gcgcccaccg gaaggagctg actgggttga aggctctcaa

Fig 13 (cont'd)

Fig 13 (cont'd)

```
4501 gggcatcggt cgagatcccg gtgcctaatg agtgagctaa cttacattaa ttgcgttgcg
4561 ctcactgccc gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca
4621 acgcgcgggg agaggcggtt tgcgtattgg gcgccagggt ggtttttctt ttcaccagtg
4681 agacgggcaa cagctgattg cccttcaccg cctggccctg agagagttgc agcaagcggt
4741 ccacgctggt ttgccccagc aggcgaaaat cctgtttgat ggtggttaac ggcgggatat
4801 aacatgagct gtcttcggta tcgtcgtatc ccactaccga gatgtccgca ccaacgcgca
4861 gcccggactc ggtaatggcg cgcattgcgc ccagcgccat ctgatcgttg gcaaccagca
4921 tcgcagtggg aacgatgccc tcattcagca tttgcatggt ttgttgaaaa ccggacatgg
4981 cactccagtc gccttcccgt tccgctatcg gctgaatttg attgcgagtg agatatttat
5041 gccagccagc cagacgcaga cgcgccgaga cagaacttaa tgggcccgct aacagcgcga
5101 tttgctggtg acccaatgcg accagatgct ccacgcccag tcgcgtaccg tcttcatggg
5161 agaaaataat actgttgatg ggtgtctggt cagagacatc aagaaataac gccggaacat
5221 tagtgcaggc agcttccaca gcaatggcat cctggtcatc cagcggatag ttaatgatca
5281 gcccactgac gcgttgcgcg agaagattgt gcaccgccgc tttacaggct tcgacgccgc
5341 ttcgttctac catcgacacc accacgctgg cacccagttg atcggcgcga gatttaatcg
5401 ccgcgacaat ttgcgacggc gcgtgcaggg ccagactgga ggtggcaacg ccaatcagca
5461 acgactgttt gcccgccagt tgttgtgcca cgcggttggg aatgtaattc agctccgcca
5521 tcgccgcttc cactttttcc cgcgttttcg cagaaacgtg gctggcctgg ttcaccacgc
5581 gggaaacggt ctgataagag acaccggcat actctgcgac atcgtataac gttactggtt
5641 tcacattcac caccctgaat tgactctctt ccgggcgcta tcatgccata ccgcgaaagg
5701 ttttgcgcca ttcgatggtg tccgggatct cgacgctctc ccttatgcga ctcctgcatt
5761 aggaagcagc ccagtagtag gttgaggccg ttgagcaccg ccgccgcaag gaatggtgca
5821 tgcaaggaga tggcgcccaa cagtcccccg gccacggggc ctgccaccat acccacgccg
5881 aaacaagcgc tcatgagccc gaagtggcga gcccgatctt ccccatcggt gatgtcggcg
5941 atataggcgc cagcaaccgc acctgtggcg ccggtgatgc cggccacgat gcgtccggcg
6001 tagaggatcg agatcgatct cgatcccgcg aaattaatac gactcactat aggggaattg
6061 tgagcggata acaattcccc tctagaaata attttgttta actttaagaa ggagatatac
6121 atatgagcga taaaattatt cacctgactg acgacagttt tgacacggat gtactcaaag
6181 cggacggggc gatcctcgtc gatttctggg cagagtggtg cggtccgtgc aaaatgatcg
6241 ccccgattct ggatgaaatc gctgacgaat atcagggcaa actgaccgtt gcaaaactga
6301 acatcgatca aaaccctggc actgcgccga aatatggcat ccgtggtatc ccgactctgc
6361 tgctgttcaa aaacggtgaa gtggcggcaa ccaaagtggg tgcactgtct aaaggtcagt
6421 tgaaagagtt cctcgacgct aacctggccg gttctggttc tggccatatg caccatcatc
6481 atcatcattc ttctggtctg gtgccacgcg gttctggtat gaaagaaacc gctgctgcta
6541 aattcgaacg ccagcacatg gacagcccag atctgggtac cgacgacgac gacaaggcca
6601 tggctgatat cggatccg
```

nucleotide sequence of pDW2700

```
   1 gatccccagc ttgcatgcct gcaggtcgag gcatttgaat tgggggtggt ggacagtaac
  61 tgtctgtaat aataattact cctgaccagg ttgcaattcg agttttgata agcataatta
 121 taccttgtac attgtgggtt ttgtgctgtg gacgttttat tgtggacatc cccataagct
 181 acaagaaacc aaaaatgaaa ttaaaagtat tgaaaaacgt cgtaacattt tatatctgag
 241 tagtatcctt tgctttaaat gtccataaaa ataatttttat aatcaataaa acaacgtttg
 301 taaatcaact gagtttacaa gtagagacat tgagggatac tttcactatg ctaaagtgaa
 361 taatcgacca aataataact cactttggta tttattcctg tcttataatg ttatgtatga
 421 attaaattca tatgcatatg gctcactctg acaaaaaaaa ataatcttcc agatcaatat
 481 tgactaccga tgcgggtggt cttttgcttt gaattctgct gaactttaca ccccgaacag
 541 caatgtgtgc ttcagctaaa aaaaagtaag tgtgttaatc agtccccccg attcttcatt
 601 ttttgccccct ctctcccgtt tcgtcggcaa aagaagagaa aataaagata agtctcaaga
 661 taggttggta atcgctaaag tggttgtgtg gataagagta gcaaaatggc aggaagagca
 721 ctttgcgcgc acacactgta ctcattgttc tggataaaat tctctcgttg tttgccgtcg
 781 gatgtctgcc tctctgccat tgagccggct tcttcactat ctttagttaa cctaaaatgc
 841 cgtttctttt ctcgtatccc actatccgtt gaggttctct gctctcttcg ctcccttacc
 901 gccagcgagc aactatccgt gggggcgcct tgctcggaag atgggggga agaaagaaga
 961 tttttgctat ttgcacttga gaaagagact tttcctgcgt cgatggttag agaacagtgt
1021 gcagacactt ttcagctacc tagatacatg gatatccccg cctcccaatc cacccaccca
1081 gggaaaaaga agggctcgcc gaaaaatcaa agttatctcc aggctcgcgc atcccaccga
1141 gcggttgact tctctccacc acttttcatt ttaaccctcg gggtacggga ttggccaaag
1201 gacccaaagg tatgtttcga atgatactaa cataacatag aacattttca ggaggaccct
1261 tggctagcgt cgacggtacc atggggcgcg ccgaattcgt taactgatca ctcgagatgc
1321 atggccggcc gagctccgca tcggccgctg tcatcagatc gccatctcgc gcccgtgcct
```

Fig 14

```
1381 ctgacttcta agtccaatta ctcttcaaca tccctacatg ctctttctcc ctgtgctccc
1441 acccccctatt tttgttatta tcaaaaaaac ttcttcttaa tttctttgtt ttttagcttc
1501 ttttaagtca cctctaacaa tgaaattgtg tagattcaaa aatagaatta attcgtaata
1561 aaaagtcgaa aaaaattgtg ctccctcccc ccattaataa taattctatc ccaaaatcta
1621 cacaatgttc tgtgtacact tcttatgttt tttttacttc tgataaattt tttttgaaac
1681 atcatagaaa aaaccgcaca caaaatacct tatcatatgt tacgtttcag tttatgaccg
1741 caatttttat ttcttcgcac gtctgggcct ctcatgacgt caaatcatgc tcatcgtgaa
1801 aaagttttgg agtatttttg gaatttttca atcaagtgaa agtttatgaa attaattttc
1861 ctgcttttgc tttttggggg tttcccctat tgtttgtcaa gagtttcgag gacggcgttt
1921 ttcttgctaa aatcacaagt attgatgagc acgatgcaag aaagatcgga agaaggtttg
1981 ggtttgaggc tcagtggaag gtgagtagaa gttgataatt tgaaagtgga gtagtgtcta
2041 tggggttttt gccttaaatg acagaatata ttcccaatat accaaacata actgtttcct
2101 actagtcggc cgtacgggcc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc
2161 tgacacatgc agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga
2221 caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc ggggctggct taactatgcg
2281 gcatcagagc agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc
2341 gtaaggagaa aataccgcat caggcggcct taagggcctc gtgatacgcc tatttttata
2401 ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc ggggaaatgt
2461 gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc cgctcatgag
2521 acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga gtattcaaca
2581 tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt ttgctcaccc
2641 agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag tgggttacat
2701 cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag aacgttttcc
2761 aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgta ttgacgccgg
2821 gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg agtactcacc
2881 agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat
2941 aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag gaccgaagga
3001 gctaaccgct tttttgcaca acatgggggа tcatgtaact cgccttgatc gttgggaacc
3061 ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg tagcaatggc
3121 aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc ggcaacaatt
3181 aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg cccttccggc
3241 tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg gtatcattgc
3301 agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtca
3361 ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca
3421 ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa aacttcattt
3481 ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca aaatcccttа
3541 acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg
3601 agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac cgctaccagc
3661 ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag
3721 cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc accacttcaa
3781 gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag tggctgctgc
3841 cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggc
3901 gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc gaacgaccta
3961 caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgcttc ccgaagggag
4021 aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca cgagggagct
4081 tccagggggа aacgcctggt atctttatag tcctgtcggg tttcgccacc tctgacttga
4141 gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc
4201 ggcctttttа cggttcctgg cctttgctg gccttttgct cacatgttct ttcctgcgtt
4261 atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata ccgctcgccg
4321 cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc gcccaatacg
4381 caaaccgcct ctccccgcgc gttggccgat tcattaatgc agctggcacg acaggtttcc
4441 cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg agttagctca ctcattaggc
4501 accccaggct ttacacttta tgcttccggc tcgtatgttg tgtggaattg tgagcggata
4561 acaatttcac acaggaaaca gctatgacca tgattacgcc aagcttgcat gcctgcaggt
4621 cgactctaga g
```

Fig 14 (cont'd)

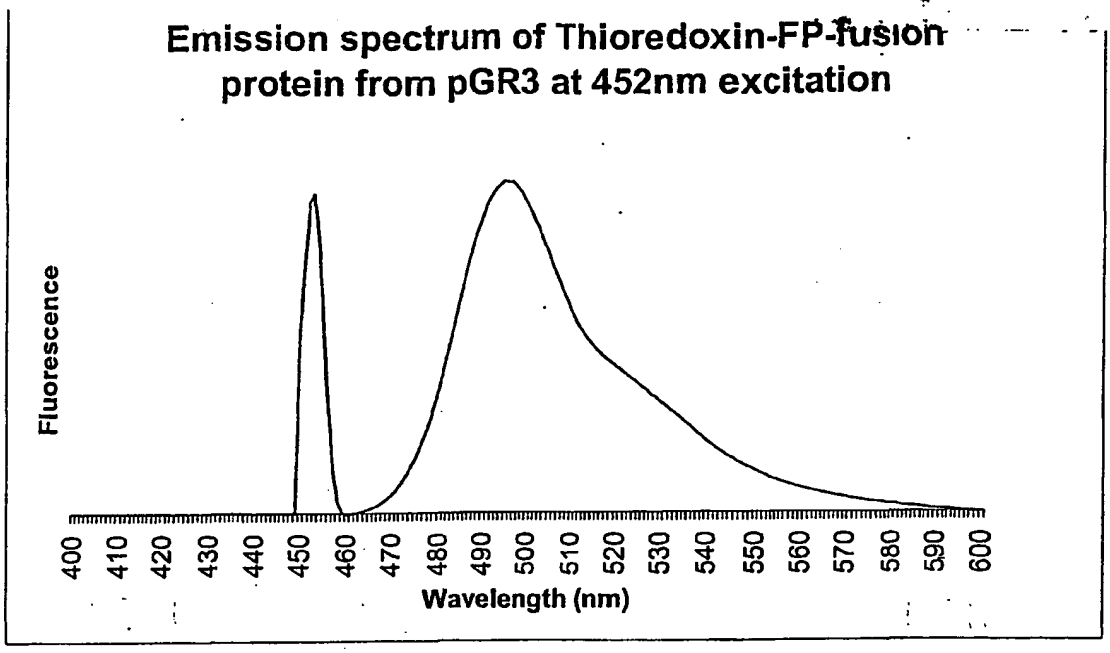


Fig 15(a)

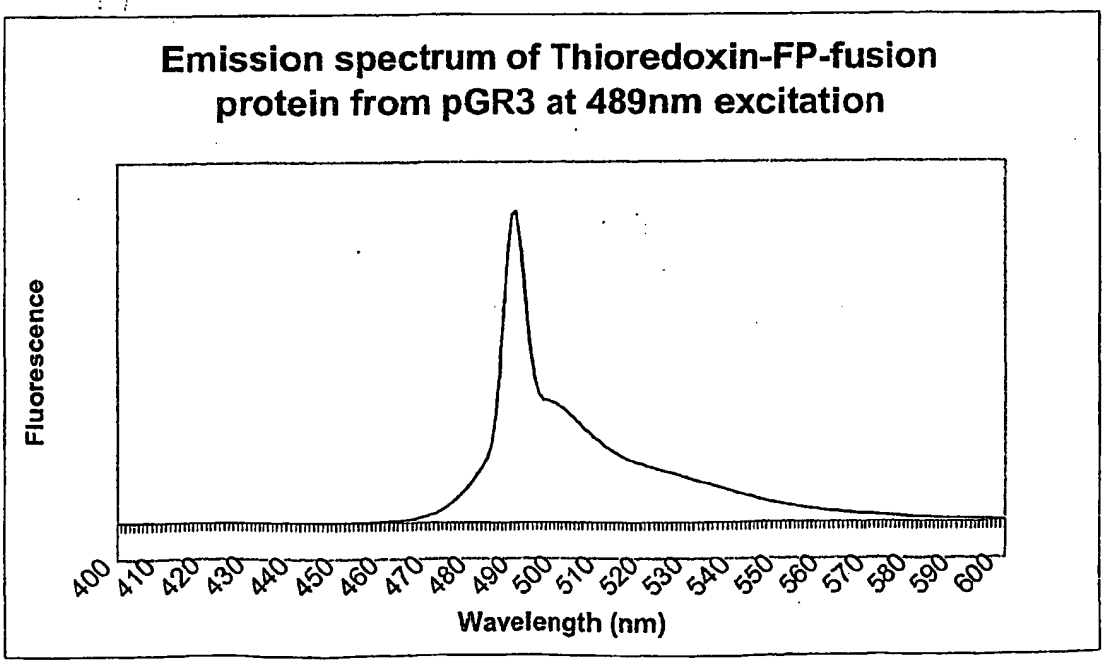


Fig 15(b)

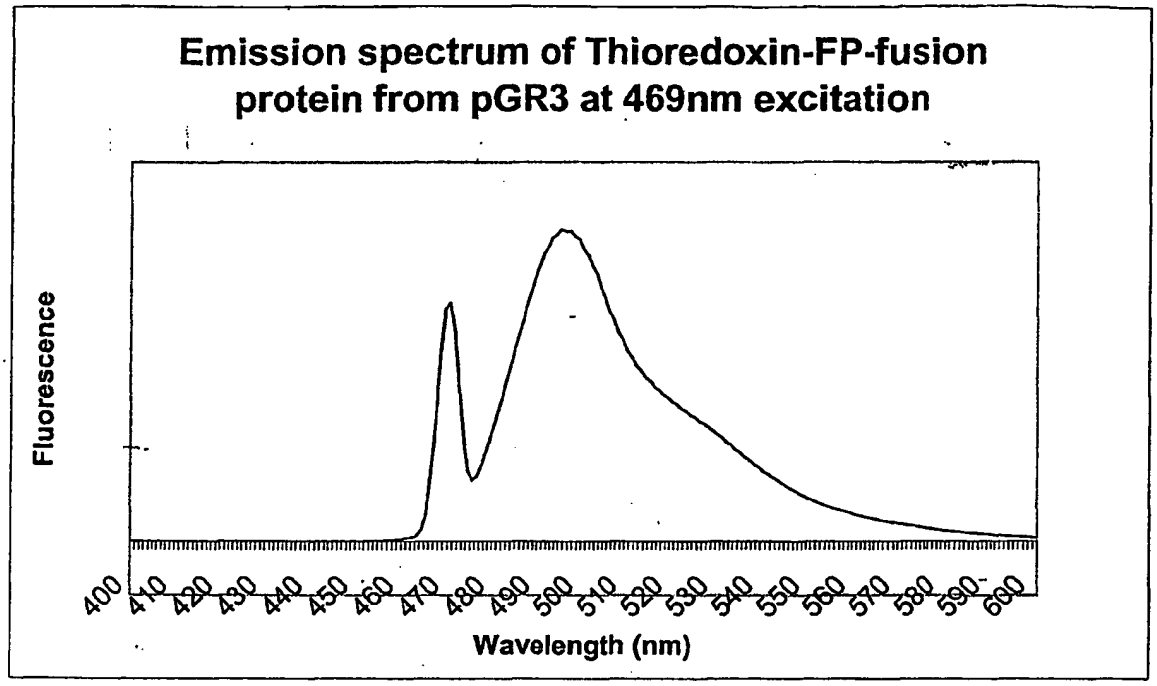
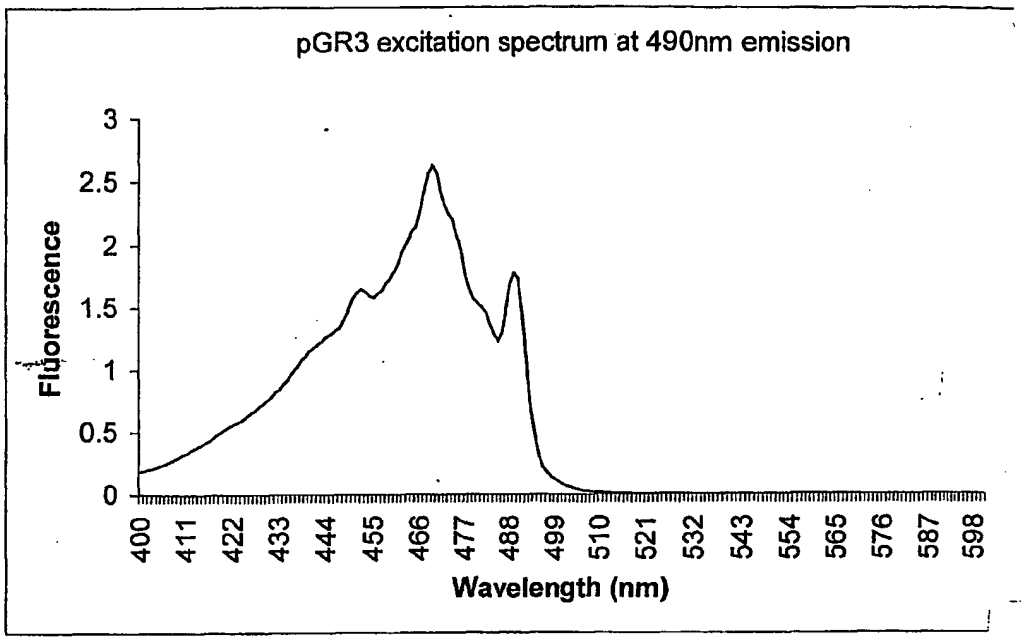

Emission spectrum of Thioredoxin-FP-fusion protein from pGR3 at 469nm excitation
Fluorescence
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
Wavelength (nm)

Fig 16

Fig 17


pGR3 excitation spectrum at 490nm emission
Fluorescence
3
2.5
2
1.5
1
0.5
0
400
411
422
433
444
455
466
477
488
499
510
521
532
543
554
565
576
587
598
Wavelength (nm)

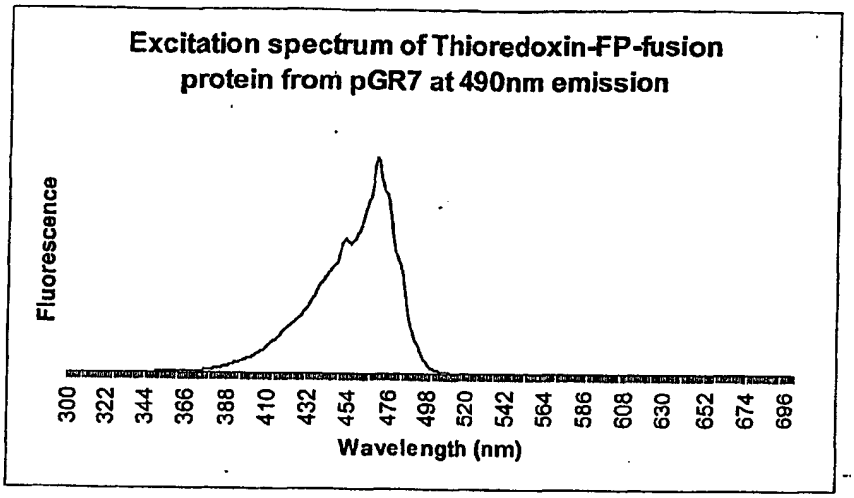

Excitation spectrum of Thioredoxin-FP-fusion protein from pGR7 at 490nm emission
Fluorescence
300
322
344
366
388
410
432
454
476
498
520
542
564
586
608
630
652
674
696
Wavelength (nm)

Fig 18

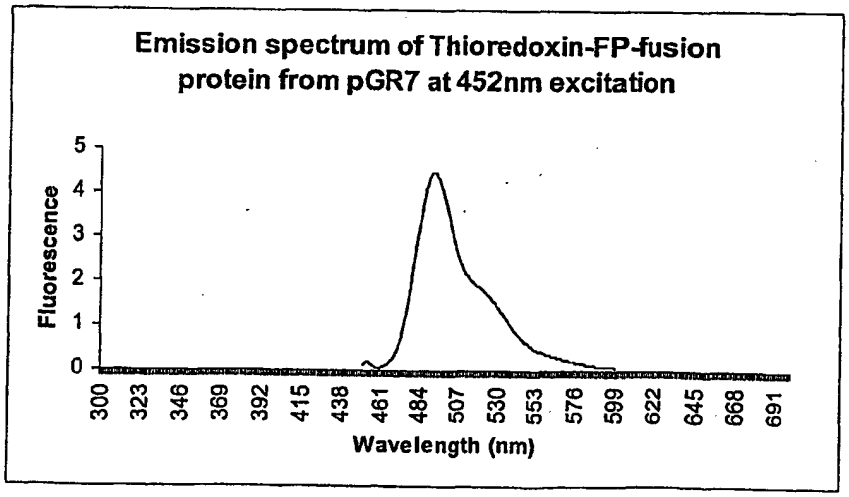

Emission spectrum of Thioredoxin-FP-fusion protein from pGR7 at 452nm excitation
Fluorescence
0
1
2
3
4
5
300
323
346
369
392
415
438
461
484
507
530
553
576
599
622
645
668
691
Wavelength (nm)

Fig 19

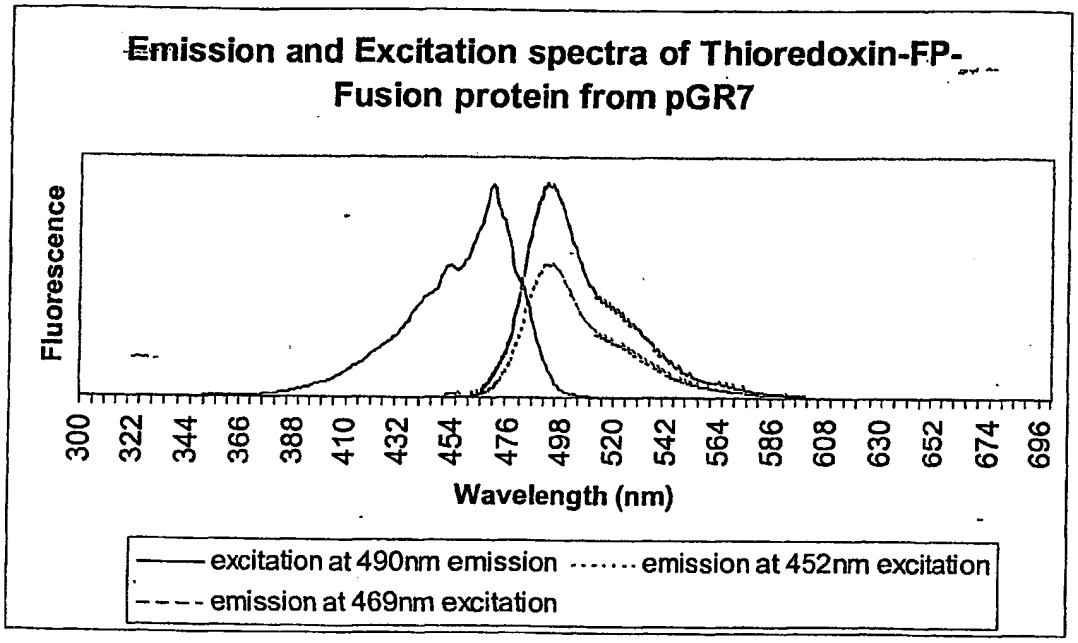

Emission and Excitation spectra of Thioredoxin-FP-
Fusion protein from pGR7
Fluorescence
300
322
344
366
388
410
432
454
476
498
520
542
564
586
608
630
652
674
696
Wavelength (nm)
excitation at 490nm emission
emission at 452nm excitation
emission at 469nm excitation

Fig 2D

Fig 21

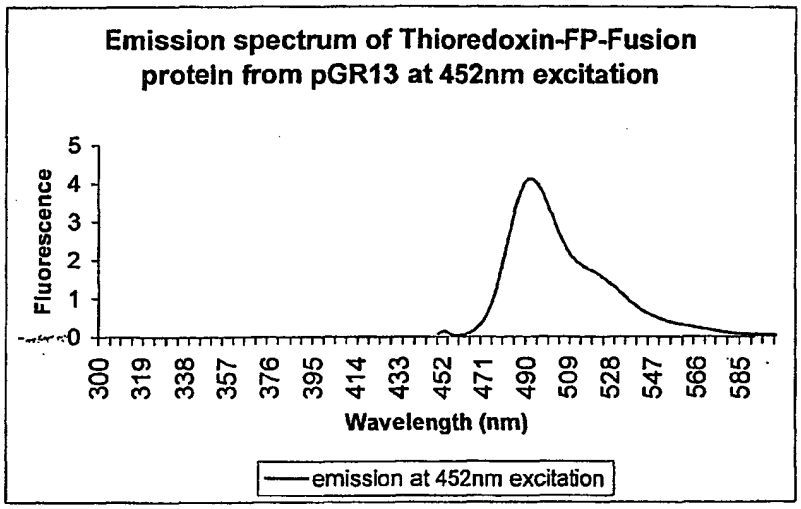
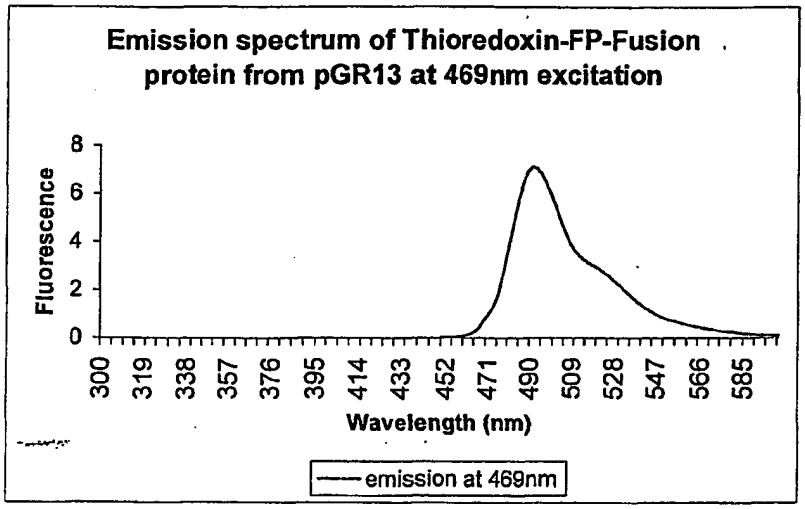

Emission spectrum of Thioredoxin-FP-Fusion protein from pGR13 at 452nm excitation
Fluorescence
0
1
2
3
4
5
300
319
338
357
376
395
414
433
452
471
490
509
528
547
566
585
Wavelength (nm)
emission at 452nm excitation

Fig 22:


Emission spectrum of Thioredoxin-FP-Fusion protein from pGR13 at 469nm excitation
Fluorescence
0
2
4
6
8
300
319
338
357
376
395
414
433
452
471
490
509
528
547
566
585
Wavelength (nm)
emission at 469nm

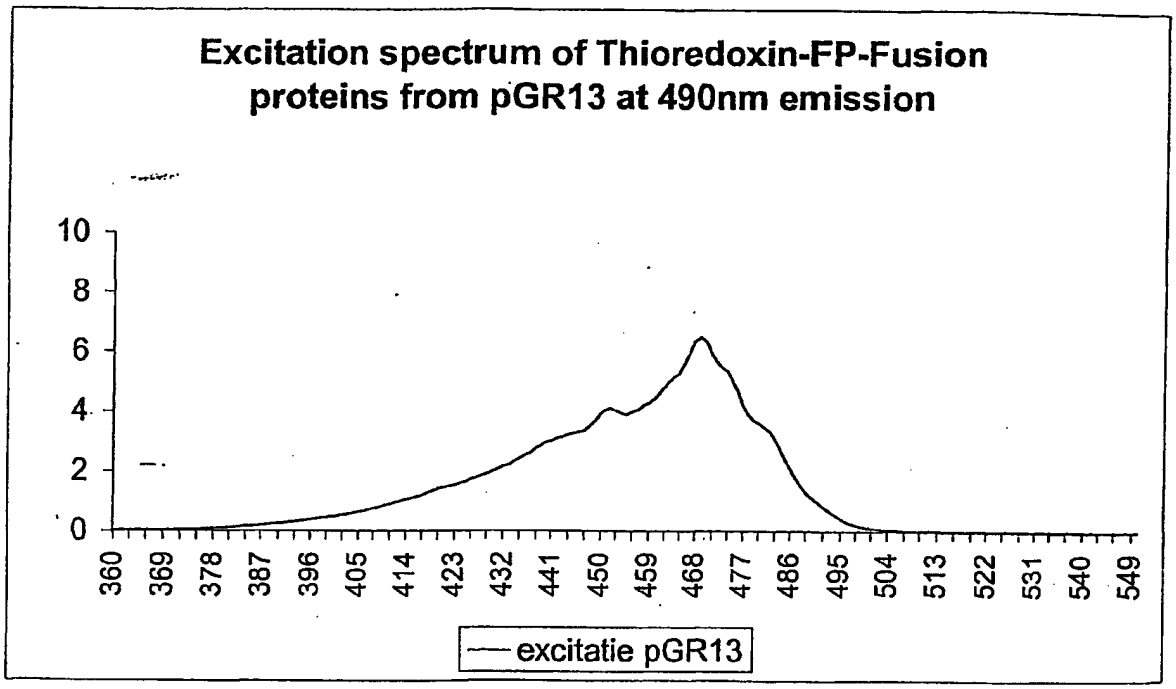
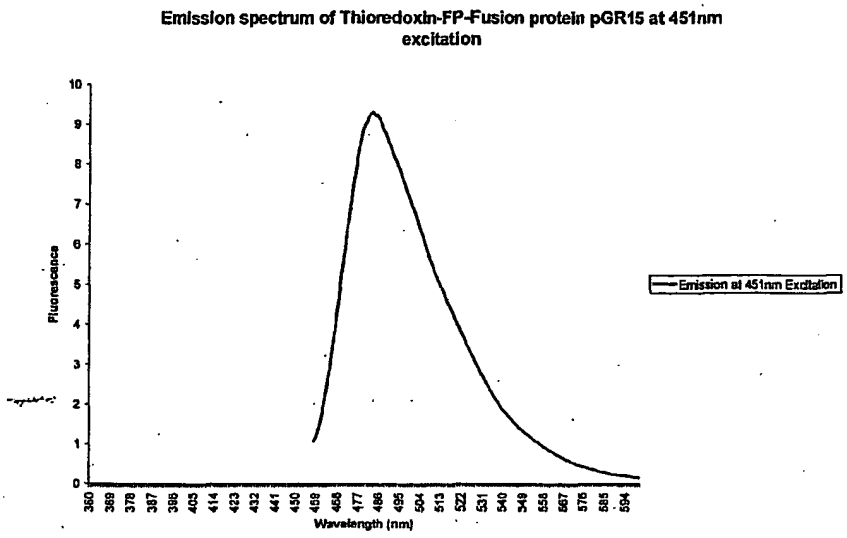

Excitation spectrum of Thioredoxin-FP-Fusion proteins from pGR13 at 490nm emission
10
8
6
4
2
0
360
369
378
387
396
405
414
423
432
441
450
459
468
477
486
495
504
513
522
531
540
549
excitatie pGR13

Fig 23

Fig 24:


Emission spectrum of Thioredoxin-FP-Fusion protein pGR15 at 451nm excitation
Fluorescence
Wavelength (nm)
Emission at 451nm Excitation

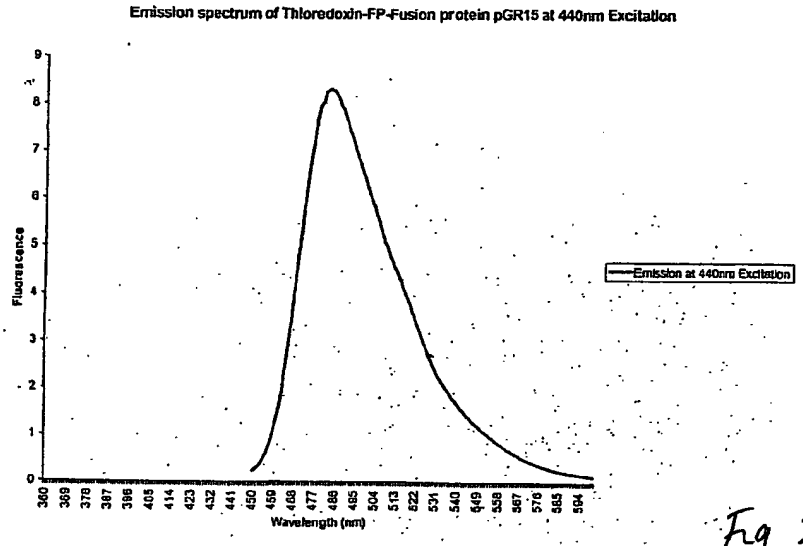
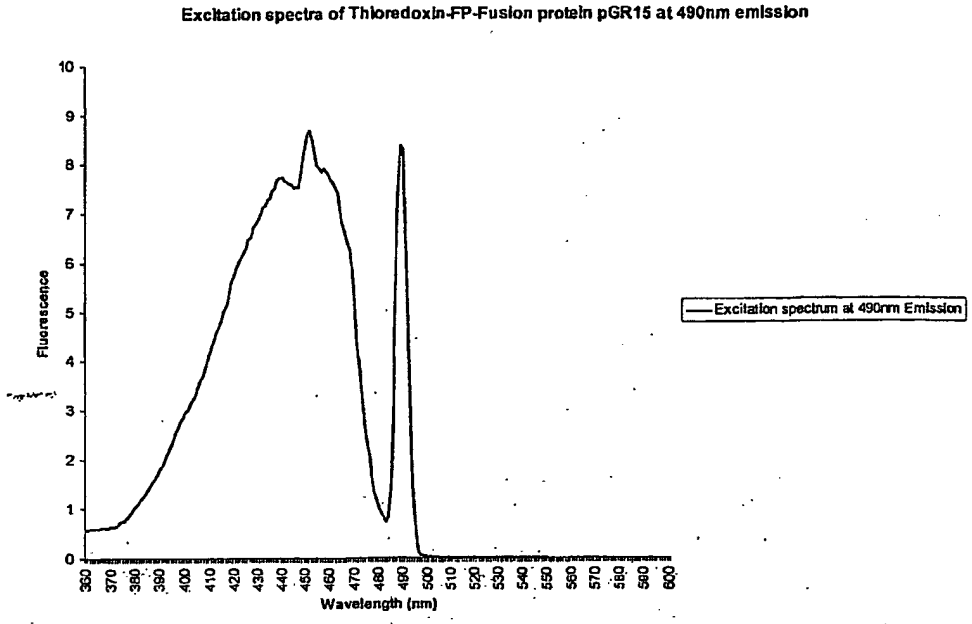

Emission spectrum of Thioredoxin-FP-Fusion protein pGR15 at 440nm Excitation
Fluorescence
Wavelength (nm)
Emission at 440nm Excitation

Fig 25

Fig 26


Excitation spectra of Thioredoxin-FP-Fusion protein pGR15 at 490nm emission
Fluorescence
Wavelength (nm)
Excitation spectrum at 490nm Emission

List of pGR clones

pGR	FP #	book	vector	Insert	mutation	What ?
pGR1	FP106	165p67, 69	pCR-XL-TOPO	fr#34: RT-PCR with oGR32/oGR34 on RNA Ca (165,p18)	non	Polythoa 2: Fluorescent colonies E.coli
pGR2	FP107	165p67, 69	pCR-XL-TOPO	fr#34: RT-PCR with oGR32/oGR34 on RNA Ca (165,p18)	non	Polythoa 2: Fluorescent colonies E. coli
pGR3	FP159	165p72, 80	pET32A-EcoRI/XhoI	fr#55: oGR69/oGR70 on pGR8	N41D, 3' end	Polythoa 2: Fluorescent colonies E. coli
pGR4	FP164	165p72, 80	pCDNA3-EcoRI/XhoI	fr#55: oGR69/oGR70 on pGR8	N41D, 3' end	Polythoa 2: Fluorescence in COS
pGR5	FP167	165p72, 80	pCDNA3-EcoRI/XhoI	fr#56: oGR69/oGR71 on pGR8	N41D, 3' end	Polythoa 2: Fluorescence in COS
pGR6	FP136	165p72, 80	pDW2721-AscI/XhoI	fr#54: oGR74/oGR72 on pGR1	non	**Polythoa 2: Fluorescence in C. elegans (and COS)**
pGR7	FP148	165p72, 79	pET32A-EcoRI/XhoI	fr#53: oGR68/oGR72 on pGR1	Q135R	Polythoa 2: Fluorescence in E. coli
pGR8	FP116	165p67, 69	pCR-XL-TOPO	fr36: RT-PCR with oGR36/oGR38 on RNA Ca (165,p18)	N41D, 3' end	Polythoa 2: : Fluorescence in in E. coli
pGR9	FP236	195p40	pDW2721-NheI/FseI	fr#68: NheI/FseI fr of pGR16 (195p38)	I106T	synthetic worm polythoa 2
pGR10	FP237	195p40	pDW2721-NheI/FseI	fr#68: NheI/FseI fr of pGR16 (195p38)	I106T	**synthetic worm polythoa 2:** : Fluorescence in C. elegans
pGR11	FP238	195p40	pDW2721-NheI/FseI	fr#73: NheI/FseI fr of FP211 (195p38)	insertion, 3 mutations	synthetic worm polythoa 2
pGR12	FP239	195p40	pDW2721-NheI/FseI	fr#74: NheI/FseI fr of FP212 (195p38)	deletion, mutation	synthetic worm polythoa 2
pGR13	FP320	195p72	pET32A-EcoRI/XhoI	Remutagenesis on pGR7 with oGR90/oGR91	non	backmutated pGR7: Polythoa 2: : Fluorescence in E. coli

Fig 27

pGR14	FP260	195p51, 53	pCR-XL-TOPO	fr#99: 525bp EcoRI/StuI fr of pGR17 in 3736bp EcoRV fr of pGR1	NONE	Hybrid-Polyth2-Disco1: : Fluorescence in E. coli
pGR15	FP337	195p80	(pGR14)FP260-HindIII/EcoRI	fr#142: 124bp fr HindIII/EcorI of pCDNA3.1/Hisa in 4212bp EcoRI/HindIII fr of pGR14	NONE	Hybrid-Polyth2-Disco1-His Tag: : Fluorescence in E. coli
pGR16	FP204	195p35	pCR-XL-TOPO	synthetic fragment Entechelon: 195p32	I106T	synthetic worm polythoa 2
pGR17	FP176		pCR-XL-TOPO	fr #45: oGR39/oGR20 op SMART cDNA #16	NONE	'3 RACE Discosoma 1
pGR18	FP326	195p67, 68	pCDNA3-EcoRI/XhoI	fr#129: 705bp EcoRI/XhoI fr of pGR19 in pcDNA3/EcoRI/XhoI	NONE	Hybrid-Polyth2-Disco1: Fluorescence in COS
pGR19	FP312	195p67	pCRblunt	fr #117: oGR69/oGR96 on pGR14	NONE	Hybrid-Polyth2-Disco1: Fluorescence in E. coli
pGR20	FP325	195p68	pDW2700-AscI/XhoI	fr#128: 700bp AscI/XhoI fr of pGR21	K221N	Hybrid-Polyth2-Disco1: : Fluorescence in C. elegans
pGR21	FP309	195p67	pCR-XL-TOPO	oGR75/oGR96 on pGR14	K221N	Hybrid-Polyth2-Disco1: Fluorescence in E. coli
pGR22	FP217	195p22 till 28	pCR-XL-TOPO	fr#65: PCR op FP58 and FP84 with oGR33, oGR34, oGR92, oGR93	Silent Mutation	Polythoa 1
pGR23	FP241	195p43	pCR-XL-TOPO	oGR84/oGR85 on pGB3202	Unknown	eGFP: fluorescent colonies
pGR24	FP327	195p70	pCDNA3-EcoRI/XhoI	fr #133: EcoRI/XhoI fr of pGR26	none	Polythoa 1: Fluorescence in COS
pGR25	FP329	195p70	pET32A-EcoRI/XhoI	fr #133: EcoRI/XhoI fr of pGR26	none	Polythoa 1
pGR26	FP317	195p70	pCR-XL-TOPO	fr #121: oGR68/oGR72 on pGR22 (195p69)	NONE	Polythoa 1
pGR						
pGR						
pGR						
pGR						

Fig 27 (cont'd)

clone pGR3:		wavelength at max fluorescense	% of max at different wavelengths
emission:			
at 452 nm excitation	Fig 15a	496nm	480nm = 30%; 525nm = 40%
at 469 nm excitation	Fig 16	496nm	480nm = 35%; 525nm = 40%
at 489 nm excitation	Fig 15b	496nm	480nm = 36%; 525nm = 40%
exitation			
at 490 nm emission	Fig 17	469nm	452nm = 64%; 456nm = 59% ; 486nm = 47%; 489nm = 68%
clone pGR7:			
emission:			
at 452 nm excitation	Fig 19	490nm	480nm = 44%; 525nm = 37%
at 469 nm excitation	Fig 20	490nm	480nm = 51%; 525nm = 38%
exitation			
at 490 nm emission	Fig 18	469nm	440nm = 53%; 452nm = 60%; 456nm = 57%; 480nm = 53%
clone pGR13:			
Emission:			
at 452 nm excitation	Fig 21	490nm	480nm = 44%; 525nm = 37%
at 469 nm excitation	Fig 22	490nm	480nm = 51%; 525nm = 38%
exitation			
at 490 nm emission	Fig 23	469nm	440nm = 53%; 452nm = 60%; 456nm = 57%; 480nm = 53%
clone pGR15:			
emission:			
at 440 nm excitation	Fig 25	484nm	470nm = 54%; 525nm = 39%
at 451 nm excitation	Fig 24	484nm	470nm = 44%; 525nm = 36%
exitation			
at 484 nm emission	Fig 26	451nm	420nm = 66%; 440nm = 91%; 447nm = 89%; 470nm = 66%

Fig 28